

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:37 ; Search time 47.68 Seconds

(without alignments)  
120.481 Million cell updates/sec

Title: US-09-407-430-1

Perfect score: 921  
Sequence: 1 MNSKGQYPTQPTYPVQPPGN.....VTORKGNFWMGSDGTYTIV 168

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	13.8	106	21	Y65155
2	116.5	12.6	1274	20	W89253
3	115.5	12.5	264	19	W71219
4	114.5	12.4	258	18	W61954
5	111.5	12.1	572	18	W31855
6	111.5	12.1	763	18	W31852
7	110	11.9	214	17	R86913
8	108.5	11.8	259	10	P91655
9	108	11.7	262	14	R42200
10	108	11.7	725	20	Y59274
11	108	11.7	1181	21	Y82707
12	108	11.7	1220	21	Y82708

13	107.5	11.7	264	12	R12531
14	107.5	11.7	278	12	R13338
15	105	11.4	114	13	R29163
16	105	11.4	259	17	R97696
17	105	11.4	1185	20	Y33497
18	104	11.3	302	20	Y02611
19	103.5	11.2	751	16	R80839
20	103.5	11.2	902	17	W02250
21	102.5	11.1	203	9	P83193
22	102.5	11.1	203	9	P83193
23	102.5	11.1	1291	20	Y01689
24	102.5	11.1	1291	20	Y01689
25	102	11.1	174	19	W81727
26	102	11.1	174	19	W81727
27	102	11.1	174	19	W81727
28	102	11.1	174	20	Y39157
29	102	11.1	174	20	Y39014
30	102	11.1	492	20	Y27246
31	101.5	11.0	626	18	W67398
32	101.5	11.0	502	20	W65730
33	101.5	11.0	1248	20	Y13464
34	101	11.0	171	13	R27476
35	101	11.0	302	20	Y02612
36	101	11.0	302	20	Y08458
37	101	11.0	302	20	Y08458
38	100.5	10.9	2441	16	R79054
39	100.5	10.9	2441	21	Y94252
40	100	10.9	303	18	W24448
41	100	10.9	303	18	W23880
42	100	10.9	303	19	W81682
43	100	10.9	303	19	W64320
44	100	10.9	303	20	Y39122
45	100	10.9	303	20	Y38985

#### ALIGNMENTS

Map2.16 expression  
Map2.16 protein incl  
PR2.1 Nicotiana a  
MBF-2 MW domain I  
Human atrophin I p  
Protein encoded by  
Japanese sea mussel  
Human transcriptio  
Bioadhesive precu  
Sequence of a bla  
Protein encoded by  
Mouse per gene pro  
M. tuberculosis im  
Mycobacterium tube  
M. tuberculosis an  
M. tuberculosis re  
C. elegans CED-6 p  
Human neuron-deriv  
NF-AT3. Mus muscu  
Human diaphanous p  
20 repeat units of  
Protein encoded by  
Protein encoded by  
Rat Plx3 protein.  
Rat Plx3 protein.  
CREB binding prote  
Mouse nuclear CREB  
Mycobacterium tube  
Mycobacterium tube  
M. tuberculosis im  
M. tuberculosis re  
M. tuberculosis re

RESULT 1  
Y65155  
Y65155 standard; Protein; 106 AA.  
AC Y65155;  
DT 01-FEB-2000 (first entry)  
DE Human 5' EST related polypeptide SEQ ID NO:1316.  
XX  
XX Human: 5' EST; expressed sequence tag; secreted protein; diagnosis;  
XX gene therapy; chromosome mapping; upstream regulatory sequence;  
XX forensic; location; development; protein synthesis; stability;  
XX regulation; identification.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W09953051-A2.  
XX  
XX 21-OCT-1999.  
XX  
XX 09-APR-1999; 99WO-IB00712.  
XX  
XX 09-APR-1998; 98US-0057719.  
XX 28-APR-1998; 98US-0069047.  
XX  
XX (GEST ) GENSET.  
XX  
XX Dumas Malne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-038446/03.  
XX N-PSDB; 242769.  
XX  
XX Novel secreted protein 5' expressed sequence tag sequences used in  
XX diagnostic, forensic, gene therapy, and chromosome mapping procedures  
XX PT

XX New nucleic acid encoding specific protein tyrosine phosphatases  
PT useful for identifying specific modulators for treatment and  
PT prevention of cancer and neurodegenerative disease  
PS  
PS Claim 2; page 160-164; 193pp; English.  
XX  
XX The present invention describes isolated, enriched or purified nucleic  
CC acids encoding PTP04, SMD, PTP05, PTP10, ALP and ALK-7 proteins. The  
CC present sequence represents human ALP. The above proteins, other than  
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
CC substances that modulate their activity (i.e. agonists and antagonists,  
CC including NMP) in vivo or in vitro. These substances are used to treat  
CC or prevent diseases associated with abnormal signal transduction  
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
CC serine/threonine kinase) are used to promote neuronal survival,  
CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
CC diseases. Nucleic acid fragments of the polynucleotides encoding the  
CC proteins can be used as probes to identify and clone related sequences/  
CC to detect protein-encoded RNA; to generate transgenic animals and in  
CC gene therapy (optionally after mutation). Ab are used to determine the  
CC proteins.  
XX  
SQ Sequence 1274 AA.

Query Match	12.6%;	Score 116.5;	DB 20;	Length 1274;
Best Local Similarity	29.4%;	Pred. No. 0.033;		
Matches 47;	Conservative 12;	Mismatches 54;	Indels 47;	Gaps 9.
QY	7 YPNQPTPYPOPCGNPVYQPTLHL-PQAP-----PYTDA-----PPAVSELTR 47			
Db	600 hpsqafgfp-qppqgplpqlqphlfpqagppllpqsgsfyyapqpyvlgqppplhtql- 657			
QY	48 PSFVHPCAA-TVPTMSAA-FPGASLILPMASVAVGLGSNIPMAXYVVGRTYPPGST 103			
Db	658 -----Pqpadqdlpahsgalrfpspppqgqphhplaygapsstrpmg-----pqaaP 704			
QY	104 VLVGGYDAGARF-----CAGATAGNIRPPPGCCP 134			
Db	705 Iltgpsasqstspshlvpsspapsqgppvprppaaep 744			
RESULT 3				
ID W71219				
W71219 standard; Protein; 264 AA.				
AC W71219;				
XX				
DT 30-OCT-1998 (first entry)				
XX				
DE Beta-D-galactoside-binding protein designated L-31-gal-lectin.				
XX				
KW Beta-D-galactoside-binding protein; L-31-gal-lectin;				
XX				
KW metastatic potential; antibody.				
XX				
OS Homo sapiens.				
XX				
OS US5801002.A.				
XX				
PN 01-SEP-1998.				
XX				
PD 22-NOV-1995; 95US-0562311.				
XX				
PF 22-NOV-1995; 94US-0188225.				
XX				
PR 26-JAN-1994; 89US-0294249.				
XX				
PR 06-JAN-1989; 91US-0681242.				
XX				
PR 05-APR-1991; 95US-0562311.				
XX				
PR 22-NOV-1995; 95US-0562311.				
XX				
PA (KARM-) KARMANOS CANCER INST BARBARA ANN.				
XX				
XX				
PI Raz A;				

XX WPI: 1998-494766/42.  
DR N-PSDB: V54736.  
XX  
XX Test for metastatic potential of cell sample - by measuring binding  
PT of antibody to L-31-gal-lectin on cell surface  
XX  
XX Disclosure: Fig 3A-B; 24pp; English.  
XX  
XX The present sequence represents a beta-D-galactoside-binding protein  
CC designated L-31-gal-lectin. The sequence is derived from clone 2. The  
CC specification describes a method for testing a cell sample for  
CC metastatic potential. The method comprises contacting the sample with  
CC a labelled antibody that binds to endogenous cell-surface  
CC L-31-gal-lectin, removing unbound antibody, and determining the amount  
CC of bound antibody as a measure of L-31-gal-lectin expression, where the  
CC metastatic potential increases as the level of L-31-gal-lectin  
CC expression increases. The antibody is produced by immunisation with a  
CC L-31-gal-lectin protein.  
XX  
SQ Sequence 264 AA:

Query Match 12.5%; Score 115.5; DB 19; Length 264;  
Best Local Similarity 32.5%; Pred. No. 0.0074;  
Matches 54; Conservative 8; Mismatches 49; Indels 55; Gaps 13;

OY 5 GQPTQP---TYPVO-PCGNPVYPTQLHLPOAPRYTDAPRAYSELYRPSFVHPCATVPT 60  
DB 42 ggyppqgqppgyppgqppg---gyp-----gqap-----psay-----pgtappgypgpt 84  
OY 61 MSAAFPGASLYLPMQASVAVGRLGSTITPMAYYPVGPITYP--PGSTVLVEGCGYDAGARFGA 118  
DB 85 apgayp-----gslapgf-----pgqpggagypspgagypgagypg 123  
OY 119 GATAGNIP---PPPGCPNMAQLAVMKG-----ANLVYTO-RKGN 155  
DB 124 psgpltpydpdpdgglmptr--mltlmglvkpnantivdlftrgn 167

RESULT 4  
ID W61954 standard; protein; 258 AA.  
XX  
AC W61954;  
XX  
DT 18-SEP-1998 (first entry)  
XX  
DE Rat galectin amino acid sequence.  
XX  
KW Mortalin; galectin; diabetes-mediating protein; insulin; DMP;  
XX diabetes; drug screening assay.  
XX  
OS Rattus sp.  
XX  
PN MO9820124-A2.  
XX  
PD 14-MAY-1998.  
XX  
PF 24-OCT-1997; 97WO-IB01627.  
XX  
PR 18-JUL-1997; 97US-0897098.  
PR 25-OCT-1996; 96US-0029324.  
PR 05-NOV-1996; 96US-0030088.  
PR 05-NOV-1996; 96US-0030186.  
XX  
PA (ANDE/) ANDERSEN H U.  
PA (CHRI/) BJERRE CHRISTENSEN U.  
PA (FEYS/) FEY S J.  
PA (KARL/) KARLSEN A E.  
PA (LARS/) MOSE LARSEN P.  
PA (NERU/) NERUP J.  
PA (POCI/) POCTOT F.

XX  
PI Andersen HU, Bjerre CHRISTENSEN U, Fey SJ, KarlSEN AE;  
PI MOSE LARSEN P, Nerup J, Poctot F;  
XX  
XX WPI: 1998-286940/25.  
DR  
XX  
XX Identification of diabetes-mediating protein(s) - by transplanting  
PT insulin-secreting cells into host at risk of developing diabetes and  
PT analysing protein expression in transplanted cells  
XX  
XX Disclosure: Fig 4; 154pp; English.  
XX  
XX This represents the amino acid sequence of murine mortalin. This is a  
CC diabetes-mediating protective protein used in the method of invention.  
CC The invention provides methods for in vivo identification of a diabetes-  
CC mediating protein (DMP) by transplanting insulin-secreting cells into  
CC host at risk of developing diabetes and analysing protein expression in  
CC transplanted cells. The DMPs are useful in drug screening assays for  
CC identifying compounds capable of modulating the development of diabetes,  
CC useful as therapeutic agents for the treatment or prevention of diabetes,  
CC and useful as targets of therapeutic agents capable of preventing or  
CC ameliorating diabetes by modulating the expression of the DMP. Changes in  
CC the expression of specific DMPs is diagnostically useful as indicator of  
CC the development of diabetes.  
XX  
SQ Sequence 258 AA:

Query Match 12.4%; Score 114.5; DB 19; Length 258;  
Best Local Similarity 31.5%; Pred. No. 0.0088;  
Matches 51; Conservative 5; Mismatches 63; Indels 43; Gaps 10;

OY 5 GQPTQP---TYPVO-PCGNPVYPTQLHLPOAPRYTDAPRAYSELYRPSFVHPCATVPT 61  
DB 33 ggyppgagypgagypgppg---yp-----gqpppsayppptpsay--pgtappgypgpt 84  
OY 62 SAAFPGASLYLPMQASVAVGRLGSTITPMAYYPVGP-IVP-PGSTVLVEGCGYDAGARFGAG 119  
DB 85 pgaifpg-----qp9g9gagypspgagypspgagypatgpfgap 120  
OY 120 ATAGNIP---PPPGCPNMAQL---AVMCGANLVYTORKN 155  
DB 121 tgppltpydpdpdgglmptrmltlmglvkpnantivdlftrgn 162

RESULT 5  
ID W31855 standard; protein; 572 AA.  
XX  
AC W31855;  
XX  
DT 27-APR-1998 (first entry)  
XX  
DE Mycobacterium tuberculosis 55 kda protein.  
XX  
KW Tuberculosis; mycobacteria; infection; diagnosis;  
XX antimycobacterial; antibiotic; vaccine.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN MO9741252-A2.  
XX  
PD 06-NOV-1997.  
XX  
PF 18-APR-1997; 97WO-EP01973.  
XX  
PR 29-APR-1996; 96DE-4017184.  
PR (GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.  
XX  
PI Espitia C, Honisch C, Moreno C, Singh M;  
DR WPI: 1997-549750/50.

RR	N-PSDB: T93610.
XX	
PT	New DNA and related proteins or RNA derived from M. tuberculosis
PT	used for diagnosis of mycobacterial infections, monitoring
PT	vaccination and development of anti-mycobacterial agents
XX	
PS	Claim 11; Fig 16; 55pp; English.
CC	This novel 55 kDa protein is encoded by an open reading frame of
CC	a Mycobacterium tuberculosis DNA fragment (see T93610) containing
CC	polymorphic GC-rich sequences. Its amino acid sequence shows
CC	a high proline content, but there is no homology to any known
CC	proline-rich antigens of mycobacteria. Novel M. tuberculosis
CC	proteins (see WJ1851-57) are claimed. These can be produced as
CC	recombinant proteins, especially in bacterial, yeast, fungal or
CC	higher eukaryote host cells, and used for diagnosing tuberculosis
CC	and other mycobacterial infections in humans or animals. The
CC	claimed proteins can also be used for epidemiological studies, for
CC	monitoring vaccination, and for the development of vaccines and
CC	anti-mycobacterial drugs.
XX	
SO	Sequence 572 AA;
Query Match	12.1%; Score 111.5; DB 18; Length 572;
Best Local Similarity	27.0%; Pred. No. 0.038; 90; Indels 59; Gaps
Matches 58; Conservative	8; Mismatches
QY	2 NSKGQVTPGTPVPVOPPGNPNVYPOTLHLQAPPYTDAPPAYSELYRP----- 48
Db	355 nppppppppptpkkl1sanppcpvpppprppappappappelppdpptppvaanspp 414
QY	49 -----SFVHPGA-----ATVPTMSAEPFGA-- 68
Db	415 appappappsalpfvnpappppttpaapksrpalpaappappbpvratttpppppppppp 474
QY	69 SYLLPMAQSVAVGPGLSGSTIPMAYPVGPPIYPGSGTVLVKGGYDAGARFGAGTAGNIIPP 128
Db	475 smalppppppipilatltpppppppplmspppppl-----ppaapdpppppitltnqpps 529
QY	129 PPGCPEPNAQLAVMOGANVLVTORKGNFEMGSDG 163
Db	530 pplapyggaaplplngfrvfark-nsligsssg 563
RESULT 6	
WJ1852	
ID	WJ1852 standard; Protein; 763 AA.
XX	
AC	WJ1852;
XX	
DT	27-APR-1998 (first entry)
DE	Mycobacterium tuberculosis 74 kDa protein.
XX	
KM	Tuberculosis; mycobacteria; infection; diagnosis;
XX	antimycobacterial; antibiotic; vaccine.
OS	Mycobacterium tuberculosis.
XX	
PN	WO9741252-A2.
XX	
PD	06-NOV-1997.
XX	
PF	18-APR-1997; 97WO-EP01973.
PR	29-APR-1996; 96DE-4017184.
PA	(GBFB ) GBF GES BIOTECH FORSCHUNG GMBH.
XX	
PI	Espilita C, Honisch C, Moreno C, Singh M;
XX	
DR	WPI; 1997-549750/50.

XX	N-PSDB; T93610.	
XX	New DNA and related proteins or RNA derived from M. tuberculosis -	
PT	used for diagnosis of mycobacterial infections, monitoring	
PT	vaccination and development of anti-mycobacterial agents	
XX		
PS	Claim 5; Fig 13; 55pp; English.	
XX		
CC	This novel 74 kDa protein is encoded by an open reading frame of	
CC	A Mycobacterium tuberculosis DNA fragment (see T93610) containing	
CC	polymorphic GC-rich sequences. Its amino acid sequence shows	
CC	a high proline content, but there is no homology to any known	
CC	proline-rich antigens of mycobacteria. Novel M. tuberculosis	
CC	proteins (see W81851-57) are claimed. These can be produced as	
CC	recombinant proteins, especially in bacterial, yeast, fungal or	
CC	higher eukaryote host cells, and used for diagnosing tuberculosis	
CC	and other mycobacterial infections in humans or animals. The	
CC	claimed proteins can also be used for epidemiological studies, for	
CC	monitoring vaccination, and for the development of vaccines and	
CC	anti-mycobacterial drugs.	
XX		
SQ	Sequence 763 AA;	
	Query Match 12.1%; Score 111.5; DB 18; Length 763;	
	Best Local Similarity 27.0%; Pred. No. 0.052;	
	Matches 58; Conservative 8; Mismatches 90; Indels 59; Gaps	
OY	2 NSKGQYPTQPIYVQPPGNVYVQTGHLPLQAPPYTDAPVASELYRP-----48	
DB	546 nspappapppprpplllsanppcpvpvpapnpppappappppelppppdprrpvpanspp 60	
OY	49 -----SFVHPGA-----ATVPYMSAARFGA-- 68	
DB	606 aappappappsalpfcvnpnpapprrpappkxrpalpaappapapayrratppppappapn 66	
OY	69 SLFLPMAOSAVAVGLSTITMATPYPGPIYPPGSGVYLVEGGYDAGARFGAGATAGNIPPP 12	
DB	666 smalpappdpplllatpppappplpmsppappl-----ppaappppaplltinqpps 72	
OY	129 PPGCPNMAOLAVMQGANVLTORKGNFFMGSSDG 163	
DB	721 pplapypgalplapllpnpgrvfarx-nslggssg 754	
RESULT 7		
R86913		
ID	R86913 standard; Protein; 214 AA.	
XX	AC R86913;	
XX	DT 10-MAY-1996 (first entry)	
XX	DE Cotton fiber-specific H6 protein.	
XX	H6 gene; promoter; cotton; transgenic plant; fiber; immobilisation;	
KW	enzyme; pesticide degradation; parathion hydrolase; extensin.	
XX	OS Gossypium hirsutum strain Coker 312.	
XX	XX	
FT	Key Location/Qualifiers	
FT	Peptide 1..25	
FT	/label= Sig-peptide	
XX		
FN	US5474925-A.	
XX		
PD	12-DEC-1995.	
XX		
PF	19-DEC-1991; 91US-0812233.	
XX		
PR	23-MAR-1994; 94US-0213327.	
PR	19-DEC-1991; 91US-0812233.	
XX		





IgE; immunoglobulin; binding protein; allergy; allergic reaction;  
absorbance; regulation; immune response.

Rattus rattus.

US5260434-A.

09-NOV-1993.

23-APR-1985; 85US-0726250.

23-APR-1985; 85US-0726250.

21-SEP-1988; 88US-0247170.

10-JUL-1991; 91US-0728125.

(SCRI) SCRIPPS RES INST.

Liu F;

WPI: 1993-367951/46.

N-PSDB: Q50878.

DNA encoding IgE-binding protein - with repetitive sequence and

homology with IgG receptor

Claim 1; Figure 1; 20pp; English.

The IgE binding protein is used to regulate or absorb IgE. It is

useful particularly in the absorbance of IgE to ease allergic

reactions.

Sequence 262 AA:

Query Match 11.7%; Score 108; DB 14; Length 262;

Best Local Similarity 30.5%; Pred. No. 0.033;

Matches 51; Conservative 8; Mismatches 58; Indels 50; Gaps 12;

5 GQYV-TOP-TYVVO-PPGNPVYPTLHPQAPYTDAPPAISELXRSFVHPCAAATVPT 60

33 GYGPSSYSGAYPGGAPPG--YP-----GQAPSAYPPTPTGSA--PQPTAPGAPPT 84

61 MSAAFGASIXILPMQSAVAVGLGSTRIMAYYPVGP-IYP--PGSTVLVEGGYDAGARFG 117

85 AQAQFPG-----GPGSPGAPPSAPGAYPSAPGAYPATGTFG 120

118 AGATAGNIP---PPPGCPNMAQLAVMOGA-----NLVTQRKGN 155

121 APTGPIVTPYDMLPGGVMPR--MLITLIGTVKPNANSILNFKKGN 165

RESULT 10

ID Y59724 standard; Protein; 725 AA.

AC Y59724;

DT 18-JAN-2000 (first entry)

DE Human normal ovarian tissue derived protein 1.

KW Human; ovary; screening; ovarian cancer; treatment.

OS Homo sapiens.

PN DEL9816395-A1.

PD 07-OCT-1999.

PF 03-APR-1998; 98DE-1016395.

PR 03-APR-1998; 98DE-1016395.

(META-) METAGEN GES GENOMFORSCHUNG MBH.

Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

WPI: 1999-552352/47.

N-PSDB: Z41222.

Nucleic acid sequences potentially useful in diagnosis or therapy of

ovarian cancer

Claim 23; Page 215; 274pp; German.

This invention describes novel nucleic acid sequences that are highly

expressed in normal ovary tissue. Artificial chromosomes and cosmid

clones containing the sequences can be used as gene transfer vehicles.

The sequences can be used to produce DNA fragments containing full-length

genes. Host cells transformed with the sequences can be used to produce

polypeptides or polypeptide fragments, which can be used to screen phage

displays for polypeptides that bind to them, or as tools for identifying

agents active against ovarian cancer, or to prepare medicaments for

treating ovarian cancer. The cDNA sequences can be used to obtain genomic

genes, their promoters, enhancers, silencers, exon structures, intron

structures and their splice variants. Y59724-Y59837 represent protein

CC fragments encoded by the cDNA sequences represented in Z41222-Z41324

CC which are derived from normal human ovarian tissue.

Sequence 725 AA:

Query Match 11.7%; Score 108; DB 20; Length 725;

Best Local Similarity 28.1%; Pred. No. 0.098; 67; Indels 52; Gaps 1;

Matches 52; Conservative 14; Mismatches 67; Indels 52; Gaps 1;

2 NSKGOYPTOPTYVOPPGNPNVYPTLHPQAPYTDAP-----AVSELYRSEVHPCAA 56

344 NAAGQPLTSPGH-----MHTQYPPYPPGPGYGPAGYPTGTSANYRQ-----QP 390

57 TVPTMSAAPP-----GASLY-----LPMQSAVAVGLGSTRIMAYYPVGPPIYPGSTVL 105

391 VAPPTSNAYPNTYISSASYSYTGSQLYAAQHQASSPTSS--PATSTP-----PPPS----- 440

106 VEGGYDAGARF---GAGA---TAGNIPPPPGCPNMAQLAVMOGANVLTQRKGNFEM 158

441 -----SGASTFGGPGGAPSSSAYALPPYTGTLPAASELPASGRTENGSDQGPML 494

159 GGSDDG 163

495 GPNG 499

RESULT 11

ID Y82707 standard; Protein; 1181 AA.

AC Y82707;

DT 11-AUG-2000 (first entry)

DE Human apoptosis related protein ABP125 SEQ ID NO.4.

KW Human; apoptosis related protein; apoptosis linked gene-2; ALC-2;

ABP125; ABP130; death trap; cancer; diagnosis; antibody.

PA (IGAK-) IGAKU SEIBUTSUGAKU KENKYUSHO KK.  
 XX WPI: 2000-353336/31.  
 DR N-PSDB: A29637.  
 XX

XX An apoptosis-related protein -  
 PS Claim 1: Page 19-23; 39pp; Japanese.  
 XX

XX The present sequence represents a human apoptosis related protein. Also  
 CC described is a method for screening a compound having activity of  
 CC inhibiting the combination between the an apoptosis related protein of  
 CC the present invention and ALG-2 (apoptosis linked gene-2) protein  
 CC including: (a) a step of contacting the apoptosis related protein with  
 CC ALG-2 protein in the presence of a compound to be tested; (b) a step of  
 CC detecting the combination between the above protein and ALG-2 protein;  
 CC and (c) a step of selecting the compound having activity of inhibiting  
 CC the combination between the apoptosis related protein and ALG-2 protein;  
 CC a compound which has activity of inhibiting the combination between the  
 CC apoptosis related protein and ALG-2 protein and can be isolated by the  
 CC above method. An antibody from the present invention and the compound  
 CC can be used for the elucidation of the molecular mechanism of apoptosis,  
 CC and for the diagnosis and treatment of diseases such as cancers.  
 CC

XX Sequence 1181 AA:

Query Match 11.7%; Score 108; DB 21; Length 1181;  
 Best Local Similarity 29.7%; Pred. No. 0.17; Mismatches 52; Gaps 10;  
 Matches 49; Conservative 12; Indels 52; Gaps 10;

OY 2 NSKQPTPTQPTPVQPCGNPVYPTQLHLPOAPPYTDAPP-----AYSELYRPSFVHPCAA 56  
 Db 813 naasqqltspgsh-----mhtqvpypypqbpqypqfgygsamytpq-----qp 859  
 OY 57 TVPTMSAAPP-----GASLY-----LPMQSVAVGPIGSTITPMAYYPVGPYPPGSTVL 105  
 Db 860 vappctsnaypntlyissassyltgsgqlyaqhgasspts--patsf--ppps--- 909  
 OY 106 VEGGYDAGARF---GAGA---TAGNIPPPPCGCPNMAOLAVMQ 143  
 Db 910 -----sgasfghgpgpappssasayalpqtgtclpaaseelpasq 948

# RESULT 12

Y82708  
 ID 182708 standard; Protein: 1220 AA.

AC Y82708;

DT 11-AUG-2000 (first entry)

DE Human apoptosis related protein ABP130 SEQ ID NO:6.

KW Human; apoptosis related protein; apoptosis linked gene-2; ALG-2;  
 KW ABP32; ABP125; ABP130; death trap; cancer; diagnosis; antibody.  
 XX

OS Homo sapiens.

PN JP2000083672-A.

PD 28-MAR-2000.

PF 11-SEP-1998; 98JP-0258165.

PR 11-SEP-1998; 98JP-0258165.

PA (IGAK-) IGAKU SEIBUTSUGAKU KENKYUSHO KK.

DR WPI: 2000-353336/31.

DR N-PSDB: A29638.

PT An apoptosis-related protein -

XX Claim 1: Page 30-35; 39pp; Japanese.  
 PS  
 XX

XX The present sequence represents a human apoptosis related protein. Also  
 CC described is a method for screening a compound having activity of  
 CC inhibiting the combination between the an apoptosis related protein of  
 CC the present invention and ALG-2 (apoptosis linked gene-2) protein  
 CC including: (a) a step of contacting the apoptosis related protein with  
 CC ALG-2 protein in the presence of a compound to be tested; (b) a step of  
 CC detecting the combination between the above protein and ALG-2 protein;  
 CC and (c) a step of selecting the compound having activity of inhibiting  
 CC the combination between the apoptosis related protein and ALG-2 protein;  
 CC a compound which has activity of inhibiting the combination between the  
 CC apoptosis related protein and ALG-2 protein and can be isolated by the  
 CC above method. An antibody from the present invention and the compound  
 CC can be used for the elucidation of the molecular mechanism of apoptosis,  
 CC and for the diagnosis and treatment of diseases such as cancers.  
 CC

XX Sequence 1220 AA:

Query Match 11.7%; Score 108; DB 21; Length 1220;  
 Best Local Similarity 29.7%; Pred. No. 0.17; Mismatches 52; Gaps 10;  
 Matches 49; Conservative 12; Indels 52; Gaps 10;

OY 2 NSKQPTPTQPTPVQPCGNPVYPTQLHLPOAPPYTDAPP-----AYSELYRPSFVHPCAA 56  
 Db 852 naasqqltspgsh-----mhtqvpypypqbpqypqfgygsamytpq-----qp 898  
 OY 57 TVPTMSAAPP-----GASLY-----LPMQSVAVGPIGSTITPMAYYPVGPYPPGSTVL 105  
 Db 899 vappctsnaypntlyissassyltgsgqlyaqhgasspts--patsf--ppps--- 948  
 OY 106 VEGGYDAGARF---GAGA---TAGNIPPPPCGCPNMAOLAVMQ 143  
 Db 949 -----sgasfghgpgpappssasayalpqtgtclpaaseelpasq 987

# RESULT 13

R12531  
 ID R12531 standard; Protein: 264 AA.

AC R12531;

DT 12-SEP-1991 (first entry)

DE Mac2.16 expression product.

KW HMEBP; leishmaniasis; Mouse Mac-2; laminin.

OS Mus musculus.

PN W09108290-A.

PD 13-JUN-1991.

PF 29-NOV-1990; 90MO-US06948.

PR 14-SEP-1990; 90US-0582628.

PR 30-NOV-1989; 89US-0444195.

PA (GEHO-) GEN HOSPITAL CORP.

DR WPI: 1991-193196/26.

DR N-PSDB: Q12207.

PT Recombinant gene encoding human macrophage carbohydrate -

PT Ige-binding protein and antibody used to treat, diagnose and  
 PT prevent e.g. inflammatory bowel disorder, leishmaniasis, hayfever  
 and bronchial



```

FT Region 78..82
FT /note= "(Pro)2-5-Gly-Tyr"
FT Region 83..87
FT /note= "(Pro)2-5-Gly-Tyr"
FT Region 88..92
FT /note= "(Pro)2-5-Gly-Tyr"
FT Region 93..99
FT /note= "(Pro)2-5-Gly-Tyr"
FT Region 100..104
FT /note= "(Pro)2-5-Gly-Tyr"
FT Region 105..109
FT /note= "(Pro)2-5-Gly-Tyr"

```

PN MO9220713-A.

PD 26-NOV-1992.

PF 15-MAY-1992: 92MO-AU00222.

PR 15-MAY-1991: 91AU-0006159.

PA (UYME ) UNIV MELBOURNE.

PI Chen C. Clarke AE;

DR MPI: 1992-415710/50.

DR N-FSDB: Q31795.

Isolated DNA sequence encoding proline rich proteins of Nicotiana glauca - used to express useful proteins and identify regulatory elements for pistil specific expression, to give e.g. longer flower life

Claim 2: Pag 47 + Fig 2; 57pp: English.

The sequence below is encoded by the 342 bp insert from clone PRP2.1. The incomplete polypeptide encoded by this cDNA contains 49% proline residues in total of 114 amino acids. It can be divided into two domains, one (encoded by nucleotides 58-210) contains 5 repeats of (Pro)4-Ala interspersed with 4 repeats of the pentapeptide Gln-Leu-Pro-Ile-Arg. These repeat sequences in turn make up two copies of a higher order repetitive unit of 23 amino acids (Pro)4-Ala-Gln-Leu-Pro-Ile-Arg-Gln-(Pro)4-Ala-Thr-Gln-Leu-Pro-Ile-Arg-Lys. The other domain (nucleotides 220-327) is composed mainly of (Pro)2-5-Gly-Tyr repeats (7) that are tandemly reiterated. At the nucleotide level, the sequences encoding the various repeats are highly conserved.

The sequences of the PRP1, PRP2, PRP3 (cDNA and genomic DNA), PRP4 genes are given in Q31794-98. The DNA sequences may be used to identify regulatory elements responsible for pistil-specific expression. Such elements may be used to provide pistil-specific expression of foreign or heterologous genes, e.g. to enable longer flower life or in the creation of male- or female-sterile plants. The PRPs may be converted by hydroxylation and glycosylation into hydroxyproline-rich glycoproteins. These could be used as effective substitutes for gums, e.g. guar gum and gum arabic, in foods and for non-food applications, where emulsification, thickening and stabilisation are required.

Sequence 114 AA:

Query Match 11.4%; Score 105; DB 13; Length 114;  
Best Local Similarity 30.8%; Pred. No. 0.024; Mismatches 45; Indels 40; Gaps 8;  
Matches 41; Conservative 7;

```

OY 8 PPGPTPVQPPGPNVPTQLHLPQAP---PYTDAPPVAVSELYRPSFVHGCATVPTMSAA 64
DB 3 PGP-----Pmqgrpppxkplippppaqlirpppatql--plrkppppaqlirqp- 54
OY 65 PFGASLYLPMASVAVGPIGSTRIPMAVYVPG---PIYPGSVLVEGCGYDAGARGAGAT 121
DB 55 -pppatqlirqp-----pppaytqppgypppypppypp-----py----- 92

```

OY 122 AGNIPPPPGCP 134  
DB 93 -----pppppypp 101

Search completed: March 6, 2001, 11:51:32  
Job time: 55 sec

Tue Mar 6 12:02:30 2001

us-09-407-430-1.rag

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:42 ; Search time 34.57 Seconds  
(without alignments)  
87.266 Million cell updates/sec

Title: US-09-407-430-1

Perfect score: 931  
Sequence: 1 MNSKQYPTQPTVPVQPGN.....VTQRKGNFMGSDGTYIW 168

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA.\*  
2: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.5	12.5	264	1	US-08-562-311-4 Sequence 4, Appl
2	114.5	12.4	574	3	US-09-275-400-6 Sequence 6, Appl
3	110	11.9	214	1	US-08-217-327-4 Sequence 4, Appl
4	108	11.7	262	3	US-08-946-914-14 Sequence 14, Appl
5	105	11.4	259	3	US-08-476-509B-29 Sequence 29, Appl
6	103.5	11.2	902	1	US-08-396-479B-6 Sequence 6, Appl
7	103.5	11.2	902	1	US-08-818-823-6 Sequence 6, Appl
8	103.5	11.1	202	5	5202236-36 Patent No. 5202236
9	102.5	11.1	203	5	5202236-1 Patent No. 5202236
10	101.5	11.0	1248	2	US-09-080-897-2 Sequence 2, Appl
11	100.5	10.9	2441	1	US-08-194-468-2 Sequence 2, Appl
12	100.5	10.9	2441	3	US-08-961-739-2 Sequence 2, Appl
13	100	10.9	1336	2	US-08-231-193A-58 Sequence 58, Appl
14	100	10.9	1336	2	US-08-486-273A-58 Sequence 58, Appl
15	100	10.9	1336	2	US-08-940-086A-58 Sequence 58, Appl
16	99.5	10.8	330	3	US-08-642-255-32 Sequence 32, Appl
17	99.5	10.8	408	1	US-07-609-716-65 Sequence 65, Appl
18	99.5	10.8	408	1	US-08-475-411A-65 Sequence 65, Appl
19	98.5	10.7	652	5	5202236-13 Patent No. 5202236
20	98.5	10.7	744	5	5202236-25 Patent No. 5202236
21	98	10.6	546	1	US-08-498-168-10 Sequence 10, Appl
22	96	10.4	677	1	US-08-188-582-13 Sequence 13, Appl
23	96	10.4	677	1	US-08-646-715-13 Sequence 13, Appl
24	96	10.4	2414	4	US-08-227-536-2 Sequence 2, Appl
25	96	10.4	2414	4	PCT-US95-04682-2 Sequence 2, Appl
26	94.5	10.3	331	5	5202236-37 Patent No. 5202236
27	94.5	10.3	331	5	5202236-3 Patent No. 5202236
28	93.5	10.2	466	3	US-08-526-136-13 Sequence 13, Appl

29	93.5	10.2	503	3	US-08-526-136-2 Sequence 2, Appl
30	93.5	10.2	505	3	US-08-526-136-4 Sequence 4, Appl
31	92.5	10.0	267	3	US-08-301-162-16 Sequence 16, Appl
32	92	10.0	1719	2	US-08-459-568-4 Sequence 4, Appl
33	92	10.0	1719	2	US-08-399-411-4 Sequence 4, Appl
34	92	10.0	1719	3	US-08-516-859A-4 Sequence 4, Appl
35	91.5	9.9	129	5	5202236-5 Patent No. 5202236
36	91.5	9.9	267	1	US-07-706-872-3 Sequence 3, Appl
37	91.5	9.9	504	3	US-09-219-849-3 Sequence 3, Appl
38	91.5	9.9	561	1	US-08-642-255-52 Sequence 52, Appl
39	91.5	9.9	720	3	US-09-219-849-4 Sequence 4, Appl
40	91.5	9.9	777	1	US-08-642-255-53 Sequence 53, Appl
41	91.5	9.9	1315	3	US-08-899-595-3 Sequence 3, Appl
42	90.5	9.8	357	1	US-07-609-716-66 Sequence 66, Appl
43	90.5	9.8	357	1	US-08-642-255-33 Sequence 33, Appl
44	90.5	9.8	357	3	US-08-475-411A-66 Sequence 66, Appl
45	90.5	9.8	1255	2	US-09-080-897-4 Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-562-311-4  
Sequence 4, Application US/08562311  
Patent No. 5801002  
GENERAL INFORMATION:  
APPLICANT: RAZ, AVRAHAM  
TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY  
OF METASTASIS IN A CELL SAMPLE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dykema Gossett  
STREET: STE 505 N. Woodward  
CITY: Bloomfield Hills  
STATE: MI  
COUNTRY: U.S.  
ZIP: 48304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/562,311  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,225  
FILING DATE:  
APPLICATION NUMBER: US 07/681,242  
FILING DATE: 04 APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/294,249  
FILING DATE: 01-JUN-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: KELLY, ROBERT L.  
REGISTRATION NUMBER: 31,843  
REFERENCE/DOCKET NUMBER: 61,686-  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 810-540-0849  
TELEFAX: 810-540-0763  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 264 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-562-311-4  
Query Match 12.5%, Score 115.5; DB 1; Length 264;  
Best Local Similarity 32.5%, Pred. No. 0.0027;





```

SEQUENCE CHARACTERISTICS:
LENGTH: 259 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

```

**Oy**    8   PTPQPTVPQAPGPNVYEQTLHLPAQPYTDAPAYSELRYPSFVHHGAATVP---TWSA    63  
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**Dd**    717   PRPPRPYSYRHPEDPAC-ETPLISEGFY-GMPPLYQTGPSPSYRRGLRMFPETRGTTC    774  
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**Oy**    64   AHPGASLILPMA-QSVAAGPLSGSTPM-----    89  
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Db 775 AQPANVFLPRPPSPDYGGGSGFPLGLPFPAPPPPLPASPPLLEGPFSQSDVHP 834  
QY 90 ----AYYPVGPPIPPGSTVLVE----GGYDAGARFGA---GATAGNIP-----PP 128  
Db 835 LPAGYKVGKVGCGGAGPDEKESRGSGSFRDVPVIGITLLEVESIIGRDLGCFPA 894  
QY 129 PPGCPNPA 136  
Db 895 PPGCEPPA 902

## RESULT 7

US-08-818-823-6  
; Sequence 6, Application US/08818823  
; Patent No. 5708158  
; GENERAL INFORMATION:  
; APPLICANT: HOEY, Timothy  
; TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; City: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,823  
; FILING DATE: 14-MAR-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/396,479  
; FILING DATE: 02-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-59450-1/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 210 277299  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 902 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-818-823-6

Query Match 11.2%; Score 103.5; DB 1; Length 902;  
Best Local Similarity 25.0%; Pred. No. 0.13;  
Matches 47; Conservative 12; Mismatches 68; Indels 61; Gaps 8;

QY 8 PTQPTYPVOPGPNVYPQTLHL-PQAPPYTDAPPAYSELIRSFVHPGATVTP-----TMSA 63  
Db 717 PPRPPYPSYPHEDPAC-ETPYLSEGFY-GMPPLYPQTPGPPSYKPLGMRPETERGTGC 774  
QY 64 AFGASLTPMA-QSVANGSLPGSTIPM----- 89  
Db 775 AQPANVFLPRPPSPDYGGGSGFPLGLPFPAPPPPLPASPPLLEGPFSQSDVHP 834  
QY 90 ----AYYPVGPPIPPGSTVLVE----GGYDAGARFGA---GATAGNIP-----PP 128  
Db 835 LPAGYKVGKVGCGGAGPDEKESRGSGSFRDVPVIGITLLEVESIIGRDLGCFPA 894  
QY 129 PPGCPNPA 136  
Db 895 PPGCEPPA 902

Db 895 PPGCEPPA 902

RESULT 8  
5202236-36  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPUA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BROADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 25-MAY-1990  
; APPLICATION NUMBER: 82,456  
; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO: 36  
; LENGTH: 202  
5202236-36

Query Match 11.1%; Score 102.5; DB 5; Length 202;  
Best Local Similarity 26.4%; Pred. No. 0.029;  
Matches 37; Conservative 9; Mismatches 59; Indels 35; Gaps 6;

QY 7 YPTQPTYPVOPGPNVYPQTLHL-PQAPPYTDAPPAYSELIRSFVHPGATVTPMSAF 65  
Db 30 YKAPSYVATYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 89  
QY 66 PG-----ASLYLPMASVAVGVLGSTITMAYVPVPIPPGSTVLVEGGYDAGARFGA 118  
Db 90 PSYPTTKPKPSYPTTKSKSLTP-SSYKPKTIYP--PTTKPKLT----- 131  
QY 119 GATAGNIP---PPGCPN 135  
Db 132 -----YPTTKPKPSYTPS 145

RESULT 9  
5202236-1  
; Patent No. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPUA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BROADHESIVE  
; PROTEIN  
; NUMBER OF SEQUENCES: 39  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/528,762  
; FILING DATE: 25-MAY-1990  
; APPLICATION NUMBER: 82,456  
; FILING DATE: 07-AUG-1987  
; APPLICATION NUMBER: 933,945  
; FILING DATE: 24-NOV-1986  
; APPLICATION NUMBER: 650,128  
; FILING DATE: 13-SEP-1984  
; SEQ ID NO: 1  
; LENGTH: 203  
5202236-1

Query Match 11.1%; Score 102.5; DB 5; Length 203;  
Best Local Similarity 26.4%; Pred. No. 0.029;  
Matches 37; Conservative 9; Mismatches 59; Indels 35; Gaps 6;

QY 7 YPTQPTYPVOPGPNVYPQTLHL-PQAPPYTDAPPAYSELIRSFVHPGATVTPMSAF 65  
Db 31 YKAPSYVATYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAK 90  
QY 66 PG-----ASLYLPMASVAVGVLGSTITMAYVPVPIPPGSTVLVEGGYDAGARFGA 118





QY 73 PMAQSVAVGELGSTITPMAYYPVGPPIYP-----PGSTVLVEGGYDAGARFGAGATAGNI 125  
DB 1034 PAAATAVAVGP--PLCRLAFDESPAPARMPRSDPESQPL--GPGAGCAGCTGGAGGCA 1089  
QY 126 PPPPPGCG---PPNAAQLAVMOGANVLTORKGNFEMGSGDGGYTTIW 168  
DB 1090 PAAPPPCFAPPPCFYLDVDSQSPS---DSRDSLSLAGASLAGLDPW 1132

DB 1090 PAAPPPCFAPPPCFYLDVDSQSPS---DSRDSLSLAGASLAGLDPW 1132

Search completed: March 6, 2001, 11:52:11  
Job time: 89 sec

## RESULT 15

US-08-940-086A-58  
; Sequence 58, Application US/08940086A  
; Patent No. 6111091  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehtman White & McAuliffe  
; STREET: 4250 Executive Square, 7th Floor  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/940,086A  
; FILING DATE: 29-SEPT-97  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231,193  
; FILING DATE: 20-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,449  
; FILING DATE: 20-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9383C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 450-8400  
; TELEFAX: (619) 450-8499  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1336 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-940-086A-58

Query Match 10.9%; Score 100; DB 3; Length 1336;  
Best Local Similarity 27.1%; Pred. No. 0.43;  
Matches 45; Conservative 9; Mismatches 74; Indels 38; Gaps 8;

QY 19 GNPVYPTQTLHPOAPPYT-----DAPPAVSELYRPSFVHPGAAVYPTMSAAPPASLYL 72  
DB 989 GRPLSPPAQPPQKPPASTFAIVRDEPEAP---PAGAFPGFSPSPAD-----1033  
QY 73 PMAQSVAVGELGSTITPMAYYPVGPPIYP-----PGSTVLVEGGYDAGARFGAGATAGNI 125  
DB 1034 PAAATAVAVGP--PLCRLAFDESPAPARMPRSDPESQPL--GPGAGCAGCTGGAGGCA 1089  
QY 126 PPPPPGCG---PPNAAQLAVMOGANVLTORKGNFEMGSGDGGYTTIW 168









[illegible]

```

C:Function:
A:Description: could act as a transcription factor antagonistic to the wg pathway
C:Keywords: DNA binding

Query Match          13.4%; Score 123.5; DB 2; Length 2715;
Best Local Similarity 31.0%; Pred. No. 0.17;
Matches 53; Conservative 4; Mismatches 81; Indels 33; Gaps 8;

QY      1 MNMKGYPPPTYPVQ-----PPGNPVYPTQLHLDPADP--YTDAAPAYSELYRPSFWHP 53
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       636 MGPGTSPGACGPPAQGPPOQPQQPPGN--YPRP---PQYPAVAATGP----- 677

QY      54 GAATPYTMSAAPGASGLYLPMAGSVAVGPLGTITMAYPVGVPIYPPGSTVLVEGGYDAG 113
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       678 ---PPPTSGAGGAGNSMPSCAGCGYPERGRPHNTGCTPPQWVPSPQOTVPCGADPG 734

QY      114 AREGAGATAGNIPPPP-PCCPMAAQLAVMGCANLV--TORKNPFMGCS 161
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       735 AAAGNHVOGCKTRPPPVVGVGPPPGSGSPRLNLKHLHKGKY--CGS 783

RESULT      8
T07796
DNA-directed RNA polymerase (EC 2.7.7.6).largest chain - soybean (fragment)
C:Species: Glycine max (soybean)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07796
R.Dietrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.
Plant Mol. Biol. 15, 207-223, 1990
A>Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A
A:Reference number: S11960; MUID:91355869
A:Accession: T07796
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-625 <DIE>
A:Cross-references: EMBL:X52492; NID:g18731; PIDN:CAA36733.1; PID:g18732
A:Experimental source: Strain var.Wayne
C:Genetics:
A:Gene: RPBI-B1
A>Note: Intron positions not resolved (Incomplete sequence)
C:Superfamily: human DNA-directed RNA polymerase II largest chain
C:Keywords: DNA binding; nucleotidylyltransferase; transcription

Query Match          13.4%; Score 123; DB 2; Length 625;
Best Local Similarity 27.7%; Pred. No. 0.039;
Matches 41; Conservative 12; Mismatches 67; Indels 28; Gaps 5;

QY      3 SKGOYPPQTPYVQPPG----NPVYPT-----LHLDPAPPYTDAPAYSELYRPS 49
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       395 SPGYSPTSPTSPSSPGTSPTSPASPTSPSPSTSPSTSPSTSPSTSPST-SPS 453

QY      50 FVHGCAATPYTMSAAFPASLYLPMAQSA-VGPLGSTRIMAYPVGVPIYPPGSVLVEG 108
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       454 YSPSTPAVSPTSPASPTSPASPTSPSPSTSPSTSPSTSPSTSPSTSPSTSP-TSP----- 508

QY      109 GTDAGARFGACATGANTDPPPPGCCPNA 136
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB       509 -----SYSPSTPAVSPTSPGYSPTS 528

RESULT      9
S14181
DNA-directed RNA polymerase (EC 2.7.7.6).largest chain (isoform B1) - soybean (fragment)
C:Species: Glycine max (soybean)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 12-Sep-1997
C:Accession: S14181
R.Dietrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.
Plant Mol. Biol. 15, 207-223, 1990
A>Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A
A:Reference number: S11960; MUID:91355869
A:Accession: S14181

```



Db 138 PPMG-----AASTGMSAGFMPPPPPPPPPPMPLNG 174

## RESULT 13

T31670  
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Mastigamoeba Invertens (frag  
C:Species: Mastigamoeba Invertens  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T31670  
R:Stillier, J.W.; Duffield, E.C.S.; Hall, B.D.  
Proc. Natl. Acad. Sci. U.S.A. 95, 11769-11774, 1998  
A:Title: Amulochondriate ameeba and the evolution of DNA-dependent RNA polymerase II.  
A:Reference number: 221052; MUID:58426228  
A:Accession: T31670  
A:Status: preliminary; translated from GR/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1690 <ST1>  
A:Cross-references: EMBL:AF083338; NID:g3511286; PID:g3511287; PIDN:MAC62246.1  
A:Experimental source: ATCC50338  
C:Genetics:  
A:Gene: RPH1  
A:Note: Intron positions not resolved (incomplete sequence)  
C:Keywords: DNA binding; nucleotidyltransferase; transcription

Query Match	12.8%;	Score 1.18;	DB 2;	Length 1690;
-------------	--------	-------------	-------	--------------

Matches 38; Conservative 13; Mismatches 71; Indels 16; Gaps 5;

0Y 5 GQYPTQPTYY----PVQPRGPNVYPQTLHLPRQAPRYTDAPRAYSELYRPSFVHPRGATVPT 60

QY 61 MSAFFPGASLYLPMAQSVAVGPIGSLTPMAIYYPVGPITYPPGSTVLVEGGYDAGAREGAG 119

[illegible]

```

      | : | | : |
Db 1614 ASPAYSPAYSPASPA 1631

```

## RESULT 14

T48627  
hypothetical protein T15N1.30 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T48627  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le  
Submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24493  
A:Accession: T48627  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-530 <EBV>  
A:Cross-references: EMBL:AL163792  
A:Experimental source: cultivar Columbia; BAC clone T15N1  
C:Genetics:  
A:Map position: 5  
A:Introns: 83/3; 176/3  
A:Note: T15N1.30

Query Match	12.88;	Score 117.5;	DB 2;	Length 530
Post [col] classification	00.00			
Post [col] classification	00.00			

Matches	51;	Conservative	9;	Mismatches	61;	Indels	55;	Gaps	9
---------	-----	--------------	----	------------	-----	--------	-----	------	---

QY 8 PTQPTYPVQPRGPNVYPQT LH---LPQAPPYTDA---PPRAYSSELYRPSFVHVGATV 58

Db 315 PQQPYQQPP---PQLQHPGYNPEEPYPQGSYPNPQRQ---PSHPPGSA-- 363

0y | : | | | |  
59 PJMSAIFPGASLY--LPMQSVAVGPlC---STIPMAVYF-----VGPIYPGCSYLVEG 108

Db 364 -----PSQQYYNAPPTPPSMYDGPGRSNSGFPSGYSPESYPTGTGPPSQYGNTPSVKP 416

QY 109 GYDAGAREGAGATAGNI PPPPGCPNAAQLAVMOGANLV TQRKGNFFMGSGDG 164

Db 417 THSGSGGAYQLPMARPLQGLPMASA-----ISSGSGGG 454

## RESULT 15

RNFE2L  
 DNA-directed RNA polymerase (EC 2.7.7.6) II 215k chain [validated] - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 04-Dec-1996 #sequence\_revision 31-Mar-1993 #text\_change 28-Jul-2000  
 C:Accession: S04457; A00693; B26777; S60151  
 R:Jokarst, R.S.; Weeks, J.R.; Zehring, W.A.; Greenleaf, A.L.  
 J Biol Chem 1994; 269:255-275. 1000

Mol. Gen. Genet. 215: 206-213, 1989  
 A:Title: Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*  
 A:Reference number: 504457; MUID:89218930  
 A:Accession: 504457  
 A:Molecule type: DNA  
 A:Residues: 1-1896 <JOK>  
 A:Cross-References: EMBL:M27431; NID:G156331; PID:AAA28868.1; PID:G156332  
 A:Biogs. J.: Seattles; L.L.; Greenleaf, A.L.

Cell 42, 611-621, 1985  
A:Title: Structure of the eukaryotic transcription apparatus: features of the gene for  
A:Reference number: A00693; MUID:55282618  
A:Accession: A00693  
A:Molecule type: DNA

A:Residues 1-318, 'GTAAY', 322-449, 'G', 451-454, 'RCTT', 459-462, 'VTGESVSS'  
A:Cross-references: EMBL:M11798  
A:Note: this sequence has been revised in reference S04457  
A:Note: the authors translated the codon AAG for residue 451 as Thr  
R:Allison, L.A.; Wong, J.K.C.; Fitzpatrick, V.D.; Moyle, M.; Ingles, C.J.

**Mod. Cell. Biol.** 8, 321-329, 1988  
**A1>Title:** The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*  
**A3:Reference number:** A93104; MUID:88094402

A: Accession: B27677  
A: Molecule type: DNA  
A: Residues: 1441-1484; 'I', 1527-1889 <ALL>

A: Cross-references: EMBL:M19537; NID:g158147; PIDN:AAA28827.1; PID:g158148  
R: Petersen, G.; Song, D.; Huebner-Doerr, B.; Oldenburg, I.; Bault, E.K.F.

Mol. Gen. Genet. 249, 425-431, 1995  
A:Title: Mapping of 11 linear epitopes recognized by monoclonal antibodies with gene-fra  
I:Reference number: S60151. MUID:96133682

A;Accession: S60151  
A;Status: preliminary; not compared with conceptual translation

```
A;Molecule type: DNA
A;Residues: 778-827 <PET>
```

C;Genetics:  
A:Gene: RPO21: BpTT215

A;Cross-references: FLYBase:FBgn0003277  
2:Map position: X 10C Y 25.7

A; Introns: 27/3; 775/3; 1526/1

```
A;Description: EC 2.7.7.6 [val
```

C;Supertfamily: human DNA-direct  
C;Keywords: DNA binding; nucle

E:67-83/Region: zinc finger CCHH motif  
E:1581-1883/Region: 7-residue repeats  
E:349/Binding site: ATP/GTP (Lys) #status predicted

Query Match 12.88; Score 117.5; DB 1; Length 1896;

Matches	40;	Conservative	10;	Mismatches	48;	Indels	33;	Gaps	7;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

QY 8 PTQPTYPVQPR--GNPVYRQTLHLRQAPRYTDAPRAY--SELYRPSF-----VHR

Db 1758 PTSPYSPSPSYDGGSPQ--YTPGSPQYSPASPKYSPTSPLYSPSSPQHSPSNQYSP 1815

QY 54 GAATV-----PTMSAEPGASLYLPMQSV-----VGPLGS-TIPMAYYPVGP 96

1816 TGSTYSATSPRYSPNMSIYSPSSTKYSPPTSPYTPPTARNYSPTSPMYSPPTAPSHYSPPTSp 1875

Tue Mar 6 12:02:32 2001

us-09-407-430-1.rpt

Page 6

QY 97 IYPPGSTVIVE 107  
| | | |  
Db 1876 AYSPSPFREE 1886

Search completed: March 6, 2001, 11:53:02  
Job time: 139 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:43 ; Search time 25.34 Seconds

(without alignments)  
214,104 Million cell updates/sec

Title: US-09-407-430-1  
Perfect score: 921  
Sequence: 1 MNSKQYPTQPTYPVPPGN.....VTQRKGNFMGSDGTYTIW 168

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues  
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	133.5	14.5	620 1	EXTN_TOBAC
2	132.5	14.4	467 1	CBPA_DICDI
3	124	13.5	1860 1	RPB0_ARATH
4	123.5	13.4	1841 1	RPB1_ARATH
5	117.5	12.8	1896 1	RPB1_DROME
6	117	12.7	1859 1	RPB1_CAEEL
7	116.5	12.6	708 1	VP40_HCVNA
8	116.5	12.6	1733 1	RPB1_YEAST
9	112	12.2	346 1	RPB1_LYCES
10	111.5	12.1	467 1	RPB1_CRIGR
11	111.5	12.1	1970 1	RPB1_HUMAN
12	111.5	12.1	1970 1	RPB1_MOUSE
13	111	12.1	497 1	HMS_DROME
14	109.5	11.9	815 1	MK07_HUMAN
15	108.5	11.8	259 1	MSP8_EIMAC
16	108.5	11.8	1324 1	IRS2_HUMAN
17	108	11.7	261 1	LEG3_RAT
18	108	11.7	1048 1	SR4_RAT
19	108	11.7	1157 1	SR4_HUMAN
20	107.5	11.7	263 1	LEG3_MOUSE
21	107.5	11.7	902 1	RPB1_DICDI
22	107	11.6	621 1	VP40_HSVBC
23	107	11.6	1752 1	RPB1_SCHPO
24	106.5	11.6	2056 1	CBP1_CAEEL
25	106	11.5	817 1	VP1_YEAST
26	105.5	11.5	875 1	FP1_MYTED
27	105.5	11.5	1083 1	T2D3_HUMAN
28	105.5	11.5	1323 1	NME4_MOUSE
29	105	11.4	1185 1	DRPL_HUMAN
30	104.5	11.3	283 1	EXTN_SORBI
31	104.5	11.3	295 1	LEG3_CANFA
32	104.5	11.3	503 1	ANXA_RABIT
33	104	11.3	268 1	CDX1_MOUSE

34	104	11.3	302 1	PIX3_MOUSE
35	104	11.3	485 1	SP62_MOUSE
36	103.5	11.2	751 1	FP1_MYTGA
37	103.5	11.2	872 1	FP1_MYTGA
38	103.5	11.2	902 1	NPC4_HUMAN
39	103.5	11.2	1183 1	DRPL_RAT
40	103	11.2	629 1	PAB2_ARATH
41	102.5	11.1	304 1	GAT1_CHICK
42	102.5	11.1	1132 1	BAT3_HUMAN
43	102.5	11.1	1291 1	PER1_MOUSE
44	102	11.1	379 1	YPS6_CAEEL
45	102	11.1	464 1	SP62_HUMAN
				035160 mus musculu
				062203 mus musculu
				027409 mytilus gal
				025434 mytilus cor
				014934 homo sapien
				P54258 rattus norv
				P42731 arabidopsi
				P17678 gallus gall
				P46379 homo sapien
				O35973 mus musculu
				O09442 caenorhabd
				O15428 homo sapien

## ALIGNMENTS

RESULT 1	ID	EXTN_TOBAC	STANDARD:	PRT:	620 AA.
AC	P13983	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	01-MAR-1992 (Rel. 21, Last annotation update)				
DE	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)				
GN	HRGNPT3				
OS	Nicotiana tabacum (Common tobacco)				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				
OC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;				
OC	Solanales; Solanaceae; Nicotiana				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. XANTHI; TISSUE=LEAF.				
RX	MEDLINE=90128263; PubMed=2612909;				
RA	Keller B., Lamb C.J.;				
RT	*Specific expression of a novel cell wall hydroxyproline-rich				
RT	glycoprotein gene in lateral root initiation."				
RL	Genes Dev. 3:1639-1646(1989).				
CC	-1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN				
CC	THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE				
CC	MAIN ROOT.				
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.				
CC	-1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE				
CC	SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN				
CC	GLYCOSYLATED.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL; X13885; CAA32090.1; -				
DR	PIR; S06733; S06733.				
KW	Repeat: Cell wall; Glycoprotein; Signal; Structural protein;				
KW	Hydroxylation.				
FT	SIGNAL	1	?		
FT	CHAIN	?	620		
FT	REPEAT	70	73		
FT	REPEAT	148	151		
FT	DOMAIN	229	242		
FT	REPEAT	229	235		
FT	REPEAT	226	242		
FT	DOMAIN	205	620		
FT	DOMAIN	499	600		
SO	SEQUENCE	620 AA;	65406 MW;		
			641DD2278AB28524 CRC64;		

Query Match 14.5%; Score 133.5; DB 1; Length 620;  
Best Local Similarity 26.2%; Pred. No. 0.013;  
Matches 48; Conservative 10; Mismatches 50; Indels 75; Gaps 7;



Query Match	13.48;	Score 123.5;	DB 1;	Length 1841.
Best Local Similarity	27.5%;	Pred. No. 0.17;		
Matches	44;	Conservative 18;	Mismatches 71;	Indels 27;
				Gaps 7;
0Y	8	PTGPTY-PVQP--PCNPVYPOT--LHRLQAPPTTDAPASSELTYRSEVHPCGAATVPTM	61	

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on legal use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M27A31 .....1

KN [3]  
X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.  
RP

[3]  
X-RAY CRYSTALLOGRAPHY (2.5 ÅNGSTROMS) OF PROTEASE.



```

RX MEDLINE-96399136; PubMed-8805707;
RA Qiu X., Culp J.S., Dillella A.G., Hellmig B., Hoog S.S., Janson C.A.,
RA Smith W.W., Abdel-Meguid S.A.;
RT "Unique fold and active site in cytomagalovirus protease.";
RL Nature 383:275-279(1996).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.
RX MEDLINE-96399137; PubMed-8805708;
RA Shieh H.-S., Kurumbail R.G., Stevens A.M., Stegeman R.A.,
RA Stutman E.J., Pak J.Y., Wiltner A.J., Palmer M.O., Wiegand R.C.,
RA "Three-dimensional structure of human cytomagalovirus protease.";
RL Nature 383:279-282(1996).
CC -1- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS A COMPONENT OF THE CAPSID
CC CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
CC ASSEMBLY IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF
CC AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
CC -1- CATALYTIC ACTIVITY: PREFERENTIALLY CLEAVES AT ALA-|-SER OR ALA-|-
CC ALA.
CC -1- PTH: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21 (SERINE PROTEASE).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17403; CA35353.1; -
DR EMBL: X17403; CA35354.1; ALT-INIT.
DR PIR: S09843; Q08EB8.
DR PDB: 1WPO; 15-OCT-97.
DR PDB: 1LAW; 26-SEP-97.
DR PDB: 1CMV; 04-SEP-97.
DR MEROPS: S21.002; -.
DR INTERPRO: IPR001847; -.
DR PFAM: PF00716; Peptidase_S21.1.
DR PRINTS: PR00236; NSVCAPSIDP40.
DR KW Coat protein; Hydrolase; Serine protease; Phosphorylation;
DR 3D-structure.
FT CHAIN 1 256 ASSEMBLIN (PROTEASE).
FT CHAIN 257 643 CAPSID ASSEMBLY PROTEIN (BY SIMILARITY).
FT PROP 644 708 C-TERMINAL PEPTIDE (BY SIMILARITY).
FT CHAIN 1 708 GENE UL80 PROTEIN.
FT CHAIN 336 708 GENE UL80.5 PROTEIN.
FT CHAIN 393 708 GENE UL80.4 PROTEIN.
FT CHAIN 478 708 GENE UL80.3 PROTEIN.
FT SITE 256 257 CLEAVAGE (BY THE PROTEASE).
FT SITE 643 644 CLEAVAGE (BY THE PROTEASE).
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
FT ACT_SITE 132 132 CHARGE RELAY SYSTEM.
FT ACT_SITE 157 157 CHARGE RELAY SYSTEM.
SQ SEQUENCE 708 AA; 73851 MW; 32A993D586824C9 CRC64;

```

Query Match 12.6%; Score 116.5; DB 1; Length 708;  
 Best Local Similarity 26.6%; Pred. No. 0.21;  
 Matches 47; Conservative 17; Mismatches 54; Indels 59; Gaps 9;

```

QY 8 PPOPTVPDPGNVYVQTLHLPQAPVYDAPVASELYRPFVHPCGATVPTMA----- 63
DB 298 PVEPSPVOPALPAPSPVLPAPESPSPSLSPSPAPL-----ASMSPHLSAAVAPPAATAPPGA 353
QY 64 -----AFPGASLYLPM-AQSAVAVCPUGSTIPMAY-----YVGP 96
DB 354 TVAGASPAVSSLAHPDGVYLPKDAFFSLGASRSVAVPMVPGAVAAPSPASPAVLP-L-P 412
QY 97 IYPP--GGSTLVLEGYDA-----GARF-GAGATAGTIPPPPECP 133
DB 413 SYPAASYGAPVY---GYDOLAAHAFADYDPRHPCGWGRYEPAPSLHPSYVPVPPPPSP 466

```

```

RESULT 8
RPNL_YEAST ID RPNL_YEAST STANDARD: PRT; 1733 AA.
AC P04050: Q12364; Q92315;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (B220).
GN RPNL OR RPN2 OR RPN220 OR SUV8 OR YDL140C OR D2150.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3644;
RX MEDLINE-85282617; PubMed-3896517;
RA Allison L.A., Moyle M., Shales M., Ingles C.J.;
RT "Extensive homology among the largest subunits of eukaryotic and
RT prokaryotic RNA polymerases.";
RL Cell 42:599-610(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE-97127826; PubMed-8972577;
RA Woelfl S., Haneman V., Saluz H.P.;
RT "Analysis of a 26,756 bp segment from the left arm of yeast
RT chromosome IV.";
RL yeast 12:1549-1554(1996).
RN [3]
RP SEQUENCE OF 1669-1733 FROM N.A.
RC STRAIN-S288C;
RX MEDLINE-95317607; PubMed-7649444;
RA Cronan J.E. Jr., Wallace J.C.;
RT "The gene encoding the biotin-apolipoprotein ligase of Saccharomyces
RT cerevisiae.";
RL FEMS Microbiol. Lett. 130:221-230(1995).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTH: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X03128; CA26904.1; -
DR EMBL: X96876; CA65619.1; -
DR EMBL: Z74188; CA98713.1; -
DR EMBL: U27182; AAC49058.1; -
DR PTR: A00692; RNB2L.
DR SGD: S0002259; RPO21.
DR INTERPRO: IPR000684; -.
DR INTERPRO: IPR00722; -.
DR INTERPRO: IPR002879; -.
DR PFAM: PF01854; RNA_pol_A2; 1.
DR PFAM: PF00623; RNA_pol_A; 1.

```





1- GENE Dev 3:1940-1956(1989).  
 CC -1- FUNCTION: ACTS AS A HOMEOTIC SELECTOR GENE CONTROLLING ANTENNAL  
 CC AND MANDIBULAR SEGMENT IDENTITY.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DEVELOPMENTAL STAGE: EMS HAS TWO DIFFERENT SPATIAL PATTERNS OF  
 CC EXPRESSION DURING EMBRYOGENESIS. THE EMS HEAD-SPECIFIC EXPRESSION  
 CC PATTERN INITIATES PRIOR TO CELLULAR BLASTODERM AND CONTINUES ONLY  
 CC UNTIL EARLY GERM-BAND EXTENSION. THE EMS METAMERIC EXPRESSION  
 CC PATTERN INITIATES AFTER GASTRULATION AND IS EXPRESSED IN THE  
 CC LATERAL NEUROBLASTS, IN ECTODERMAL CELLS AT THE ANTERIOR LATERAL  
 CC BORDERS OF EACH SEGMENT, AND IN THE FLITZOEPPER PRIMORDIA.  
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN CANTON S.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X66270; CAA46985.1; -  
 CC EMBL: X51653; CAA35965.1; -  
 CC DR PIR: A34218; A34218.  
 CC DR PIR: S22708; S22708.  
 CC DR HSSP: P06601; 1FJL.  
 CC DR TRANSFAC: T02008; -  
 CC DR FLYBASE: FBgn0000576; ems.  
 CC DR INTERPRO: IPR001356; -  
 CC DR PFAM: PF00046; homeobox; 1.

[illegible]

```

Query Match          12.1%; Score 111; DB 1; Length 497;
Best Local Similarity 26.1%; Pred. No. 0.36;
Matches 46; Conservative 14; Mismatches 60; Indels 56; Gaps 10

QY      11 PTYQPPRGNP---VYPDQ-----LHLQADPP---YTDAPPAYSELKRSFYVHP 53
          | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
DB      197 PAQRQRPHPRPKKVSQSSQSSPFLTILSSNHAIRPQOQOQDPPPNYP---KRAMNHP 253
          | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY      54 GAATVPYNSAALPCASLYLPMASVANGPLSGSTIPMA-----YYPVGGIYPP- 100
          | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
DB      254 GGAG-PMMPMPMPAGAGVRRP---PRPGGPRMPGCGQLDIALPRTYINAPPELLRPQ 308
          | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY      101 -----GSTVLVEGGYDAGARFGAGATAGNTP-----PRPGGCPNAALQ 139
          | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
DB      309 HNPFLIAAQQMAAALQAGHIVLCPRAAAAAAAG-LRPNAGQEMRNPGMARDSTYL 363
          | : : : | : : | : : | : : | : : | : : | : : | : : | : : |

RESULT 14
MK07_HUMAN          STANDARD:          PRT;          815 AA.

ID      MK07_HUMAN          Q13164; Q16634;
AC      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      30-MAY-2000 (Rel. 35, Last annotation update)
DE      MINGEN-ACTIVATED PROTEIN KINASE 7 (EC 2.7.1.-) (EXTRACELLULAR SIGNAL-
DE      REGULATED KINASE 5) (ERK5) (ERK4) (BMK1 KINASE).
GN      MAPK7 OR PRKM7 OR ERK5 OR ERK4.

```

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RETAL BRAIN;  
 RX MEDLINE=95279403; PubMed=7759517;  
 RA Zhou G., Bao Z.Q., Dixon J.E.;  
 RT "Components of a new human protein kinase signal transduction  
 pathway";  
 RL J. Biol. Chem. 270:12665-12669(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PLACENTA;  
 RX MEDLINE=95374539; PubMed=7646528;  
 RA Lee J.-D., Ulevitch R.J., Han J.;  
 RT "Primary structure of BMK1, a new mammalian map kinase";  
 RL Biochem. Biophys. Res. Commun. 213:715-724(1995).  
 CC -1- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND  
 CC NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.  
 CC -1- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE  
 CC PHOSPHORYLATION (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN  
 CC HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE  
 CC IN LIVER.  
 CC -1- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN  
 CC TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.  
 CC -1- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN  
 CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY  
 CC ROLE, IS ABSENT.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC MAP KINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U25278; AAA81381.1; -  
 DR EMBL: U29725; AAA82931.1; -  
 DR EMBL: U29726; AAA82932.1; -  
 DR EMBL: U29727; AAA82933.1; -  
 DR HSP: P24941; IAO1.  
 DR MIM: 602521; -  
 DR INTERPRO: IPR000719; -  
 DR INTERPRO: IPR002290; -  
 DR PFAM: PF00069; PKINase; 1.  
 DR PROSITE: PS01351; MAPK; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR Transferrase: Serine/threonine-protein kinase: ATP-binding; Cell cycle;  
 KW Phosphorylation.  
 FT DOMAIN 54 346  
 FT NP\_BIND 60 68  
 FT BINDING 83 83  
 FT ACT\_SITE 181 181  
 FT MOD\_RES 218 218  
 FT MOD\_RES 220 220  
 FT MOD\_RES 220 220  
 FT DOMAIN 337 340  
 FT DOMAIN 433 464  
 FT DOMAIN 520 523  
 FT DOMAIN 577 699  
 FT CONFLICT 19 32  
 FT CONFLICT 609 609  
 FT SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;  
 L -> V (IN REF. 2).  
 AREGTRPHRLCS -> GPVKEPAPHTAASVA (IN  
 REF. 2).

Query Match 11.9%; Score 109.5; DB 1; Length 815;  
 Best Local Similarity 28.9%; Pred. No. 0.73;  
 Matches 52; Conservative 17; Mismatches 56; Indels 55; Gaps 12;  
 OY 8 PTPPTVPOPG---NPVPTLHLQAAPPYTPADPAVSLYRPSVHGCATVPTMSAN 64  
 DB 598 PVQPTSP--PPGLAQTGPQ---PQSAQSTSGP---VPQACPDPGPAPHPPT--- 642  
 OY 65 FPGASLYLPMASVAVGPGSLTIPMAVYPVGPITPGSTVLVEGYDAGARFGAGATDAGN 124  
 DB 643 GPPGPVPPAPAPQIAT---STSLAAGSLVP--PPG-----LPGSSTPGV 662  
 OY 125 IPPPPPGCP---MAQLAVMGANV-LVTORKN-----FFMG---GSDGQYIT 167  
 DB 683 LPYFPLPLPPDAGAPAPSSMSSESPVNILVTQLSKSQVEDPLPVFSGHPKSGAGYGV 742

RESULT 15  
 MSP8\_ETMAC STANDARD; PRT; 259 AA.  
 AC P09125;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DE 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE MERZOITE SURFACE PROTEIN CMZ-8 (FRAGMENT).  
 OS Eimeria acervulina.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriidae; Eimeriidae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12;  
 RX MEDLINE=89041578; PubMed=3186453;  
 RA Jenkins M.C.;  
 RT "A cDNA encoding a merozoite surface protein of the protozoan Eimeria  
 RT acervulina contains tandem-repeated sequences";  
 RL Nucleic Acids Res. 16:9863-9863(1988).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X12445; CAA30977.1; -  
 DR PIR: S01704; S01704.  
 KW Merzoite; Repeat.  
 FT FT 1 1  
 FT DOMAIN 5 109  
 FT REPEAT 5 11  
 FT REPEAT 12 18  
 FT REPEAT 19 25  
 FT REPEAT 26 32  
 FT REPEAT 33 39  
 FT REPEAT 40 46  
 FT REPEAT 47 53  
 FT REPEAT 54 60  
 FT REPEAT 61 67  
 FT REPEAT 68 74  
 FT REPEAT 75 81  
 FT REPEAT 82 88  
 FT REPEAT 89 95  
 FT REPEAT 96 102  
 FT REPEAT 103 109  
 FT SEQUENCE 259 AA; 26988 MW; EFC8B014ACF40502 CRC64;  
 15 X 7 AA REPEATS OF S-P-P-S-T-P-V.  
 Query Match 11.8%; Score 108.5; DB 1; Length 259;  
 Best Local Similarity 26.2%; Pred. No. 0.29;  
 Matches 38; Conservative 18; Mismatches 58; Indels 31; Gaps 6;

```

QY 11 PTYVQPPGNPVYQOTLHPQAPPYTDAPPAXSELYRPSF-VHFGATV--PTMSAAPG 67
   | : | | | | | : | : | | | : | : | : | : | : | : | : | : | : |
Db 7 PSTPVSPSTPVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 64
QY 68 ASLYLPMAQSVAVGPIGSTIPMAYYPVGPPIYPGSTVLEGGYDAGAFGAGATAGNIPP 127
   | : | | | | | : | : | | | : | : | : | : | : | : | : | : | : |
Db 65 TPVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 98
   | : | | | | | : | : | | | : | : | : | : | : | : | : | : | : |
QY 128 PPPGCCPPNNAQLAVMAGANLVATOR 152
   | : | | | | | : | : | | | : | : | : | : | : | : | : | : | : |
Db 99 STPVSPSPSPAPAGAVGVNSSLISOR 123
   | : | | | | | : | : | | | : | : | : | : | : | : | : | : | : |

```

Search completed: March 6, 2001, 11:53:31  
Job time: 168 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:51:32 ; Search time 47.68 seconds

(Without alignments)  
86.058 Million cell updates/sec

Title: US-09-407-430-1\_COPY\_1\_120

Perfect score: 653

Sequence: 1 MNSKGOYPTQPTYPVQPCGN.....GSTVLEGGDAGARFGACA 120

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: A.Geneseq\_36.\*  
2: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	19.4	106	21 Y65155	Human 5' EST relat
2	101	15.5	1274	20 W69253	Human ALP. Homo s
3	100	15.3	171	13 R27476	20 repeat units of
4	99.5	15.2	216	9 P83192	Sequence of the BI
5	99.5	15.2	751	16 R80839	Japanese sea mussel
6	98.5	15.1	652	9 P82974	Bioadhesive precu
7	98.5	15.1	744	9 P82975	Bioadhesive precu
8	98	15.0	246	19 W61955	Human galectin ami
9	98	15.0	264	19 W71219	Beta-D-galactoside
10	98	15.0	492	20 Y27246	C. elegans CBD-6 p
11	96.5	14.8	203	9 P82970	Bioadhesive precu
12	96.5	14.8	203	9 P83193	Sequence of a blosa

13	96.5	14.8	258	19 W61954	Rat galectin amino
14	96.5	14.8	262	14 R42200	IGE binding protei
15	95	14.5	650	21 R85180	Cellulose synthase
16	94	14.4	259	17 R97696	WB-2 Wn domain 11
17	94	14.4	902	17 W02250	Human transcriptio
18	94	14.4	1051	20 Y29321	Human btaxin-2 lik
19	93.5	14.3	112	21 R84457	Amino acid sequenc
20	93.5	14.3	334	9 P82971	Bioadhesive precu
21	93.5	14.3	334	9 P83194	Sequence of a blosa
22	93.5	14.3	2783	13 R23962	AFP-1. Homo sapie
23	93.5	14.3	2783	13 R23963	AFP-1 (A1A 2460 V
24	93	14.2	214	17 R86913	Cotton fiber-speci
25	92	14.1	902	20 W65730	Sequence encoded b
26	91.5	14.0	267	14 R40171	Annexin XI type I
27	90.5	13.9	503	14 R34127	Human apoptosis re
28	90.5	13.9	505	14 R34128	Mac2.16 expression
29	90.5	13.9	1181	21 Y82707	Secreted protein 1
30	90.5	13.9	1220	21 Y82708	A human small prol
31	90	13.8	264	12 R12531	Mycobacterium tube
32	90	13.8	278	12 R13338	Mouse alpha-1 (XVI
33	89.5	13.7	137	20 Y59678	Human secreted pro
34	89.5	13.7	137	21 Y84606	Beta-D-galactoside
35	89.5	13.7	572	18 W31855	Elmeria cell surf
36	89.5	13.7	763	18 W31852	Human leukocyte ce
37	89.5	13.7	1288	18 W26326	Recombinant human
38	89.5	13.7	1288	20 W22297	Human neuron-deriv
39	89	13.6	155	19 W5148	
40	89	13.6	197	21 Y66735	
41	89	13.6	250	19 W71218	
42	88	13.5	259	10 P91655	
43	88	13.5	371	19 W64535	
44	88	13.5	466	11 R07084	
45	88	13.5	626	18 W16398	

## ALIGNMENTS

RESULT 1	Y65155	standard; Protein: 106 AA.
ID	Y65155	
XX	Y65155:	
AC	01-FEB-2000	(first entry)
XX		
DE	Human 5' EST related polypeptide SEQ ID: NO:1316.	
XX		
KW	Human: 5' EST: expressed sequence tag; secreted protein; diagnosis;	
KW	gene therapy; chromosome mapping; upstream regulatory sequence;	
KW	forensic; location; development; protein synthesis; stability;	
KW	regulation; identification.	
XX		
OS	Homo sapiens.	
XX		
PN	W09953051-A2.	
PD	21-OCT-1999.	
XX		
PF	09-APR-1999;	99WO-IB00712.
XX		
PR	09-APR-1998;	98US-0057719.
PR	28-APR-1998;	98US-0069047.
XX		
PA	(GEST ) GENSET.	
XX		
PI	Dumas Milne Edwards J, Duclert A, Giordano J;	
XX		
DR	WPI: 2000-038446/03.	
DR	N-PSDB: 242769.	
XX		
PT	Novel secreted protein 5' expressed sequence tag sequences used in	
	diagnostic, forensic, gene therapy, and chromosome mapping procedures	







CC The present sequence is that of a strongly adhesive protein isolated  
CC from *Mytilus galloprovincialis* (Japanese sea mussel). A cDNA  
CC sequence coding for the protein was identified using the probe  
CC sequence in Q99775.

SO Sequence 751 AA;

Query Match 15.2%; Score 99.5; DB 16; Length 751;  
Best Local Similarity 31.4%; Pred. No. 0.16;  
Matches 32; Conservative 6; Mismatches 43; Indels 21; Gaps 5;

QY 3 SKGQPTPTVQPPGPNVYPTQTLHL-PQAPPTDAPPAVSELY--RPSF-----VHPG 54  
Db 641 akpsyp--ptykakpsypatyptykakpsypptykakpsypptykakpsyp-----ply 693

QY 62 SAAPFGASLYLPMAQSAVAVGFLGSTIPMAYP---YVGPITPP 100  
Db 694 kskssypsykpk-----ktyppptykpkltpplykpk 725

RESULT 6

P82974  
ID P82974 standard; protein; 652 AA.

AC P82974;

DT 03-DEC-1990 (first entry)

DE Bioadhesive precursor protein from cDNA N1.

KW Bioadhesive precursor protein; cDNA N1; mussels; barnacles; oysters;

KM biomedical adhesive; sealants; wound healing.

OS *Mytilus edulis*.

PN W08803953-A.

PD 02-JUN-1988.

PF 24-NOV-1987; 87WO-US03048.

PR 07-AUG-1987; 87US-0082456.

PA (GENE-) GENEX CORP.

PI Maugh KJ, Anderson DM, Strausberg R, Strausberg SL, McCandliss R;

DR WPI; 1988-161622/23.

DR N-PSDB; N80233.

PT DNA coding for bio-adhesive precursor protein - obtd. from mussels,  
PT barnacles and oysters and used esp. to prepare biomedical adhesive.

PS Disclosure; ; pp: English.

CC mRNA from the phenol gland of the foot of *M. edulis* is isolated for  
CC the synthesis of cDNA. The cDNA is inserted into a bacteriophage  
CC and *E. coli* transfected and cultured. Five clones were isolated and  
CC characterized, one of which is clone N1.

CC Clone N1 does not have a perfect overlap with any of clones 14-1,  
CC 52, 55 or 56. Clone N1 also does not have a characteristic 5' end and  
CC therefore does not include the N-terminus of the natural gene. It does  
CC code for a polypeptide with a MW 80 kD, a substantial portion of the  
CC natural polypeptide (MW 110-130 kD).

CC The protein is chemically or enzymatically hydroxylated to produce  
CC adhesive compns. having excellent properties in wet environments. The  
CC product is partic. useful as a biomedical adhesive or sealant e. g. in  
CC wound healing or as a marine or dental adhesive. It can be used as a  
CC thin film membrane.  
CC See also N80228-34.

SO Sequence 652 AA;

Query Match 15.1%; Score 98.5; DB 9; Length 652;  
Best Local Similarity 28.4%; Pred. No. 0.17;  
Matches 31; Conservative 11; Mismatches 56; Indels 11; Gaps 4;

QY 3 SKGQPTPTVQPPGPNVYPTQTLHL-PQAPPTDAPPAVSELY--RPSF-----VHPG 54  
Db 301 akptykakpsypptykakpsypptykakpsypptykakpsypptykakpsypkpt 360

QY 55 AATVPTMSAAPPASLYLPMAQSAVAVGFLGSTIPMAY---YVGPITPP 100  
Db 361 yakaptypstykakpsypptykakpsypptykakpsypptykakpsyp 409

RESULT 7

P82975  
ID P82975 standard; protein; 744 AA.

AC P82975;

DT 03-DEC-1990 (first entry)

DE Bioadhesive precursor protein from the mussel *Mytilus edulis*.

KW Bioadhesive precursor protein; mussels; barnacles; oysters;

KM biomedical adhesive; sealants; wound healing.

OS *Mytilus edulis*.

PN W08803953-A.

PD 02-JUN-1988.

PF 24-NOV-1987; 87WO-US03048.

PR 07-AUG-1987; 87US-0082456.

PA (GENE-) GENEX CORP.

PI Maugh KJ, Anderson DM, Strausberg R, Strausberg SL, McCandliss R;

DR WPI; 1988-161622/23.

DR N-PSDB; N80234.

PT DNA coding for bio-adhesive precursor protein - obtd. from mussels,  
PT barnacles and oysters and used esp. to prepare biomedical adhesive.

PS Claim 5; Page 67+68; 233pp; English.

CC The protein is chemically or enzymatically hydroxylated to produce  
CC adhesive compns. having excellent properties in wet environments. The  
CC product is partic. useful as a biomedical adhesive or sealant e. g. in  
CC wound healing or as a marine or dental adhesive. It can be used as a  
CC thin film membrane.  
CC See also N80228-34.

SO Sequence 744 AA;

Query Match 15.1%; Score 98.5; DB 9; Length 744;  
Best Local Similarity 28.4%; Pred. No. 0.19;  
Matches 31; Conservative 11; Mismatches 56; Indels 11; Gaps 4;

QY 3 SKGQPTPTVQPPGPNVYPTQTLHL-PQAPPTDAPPAVSELY--RPSF-----VHPG 54  
Db 314 akptykakpsypptykakpsypptykakpsypptykakpsypptykakpsypkpt 373

QY 55 AATVPTMSAAPPASLYLPMAQSAVAVGFLGSTIPMAY---YVGPITPP 100  
Db 374 yakaptypstykakpsypptykakpsypptykakpsypptykakpsyp 422



XX	Y27246;
AC	08-OCT-1999 (first entry)
DT	C. elegans CED-6 protein.
XX	
DE	
XX	CED-6; phagocytosis; C. elegans; human; h1CED-6; h2CED-6; h3CED-6;
KW	signal transduction pathway; apoptotic cell; inflammation; cancer;
KW	autoimmune disease; neurodegenerative disease; sickle cell anemia;
KM	Huntington's disease; Alzheimer's disease; myocardial heart failure;
OS	amyotrophic lateral sclerosis; stroke; AIDS.
XX	
XX	Caenorhabditis elegans.
FH	Location/Qualifiers
FT	Key 49..193
FT	/note= "specifically claimed phosphotyrosine binding
FT	(PTB) domain (Y27247)"
FT	Region 194..239
FT	/note= "charged region"
FT	242..338
FT	/note= "specifically claimed proline/serine
FT	rich region (Y27248)"
PN	WO9937770-A1.
PD	29-JUL-1999.
PF	21-JAN-1999; 99WO-US01361.
XX	
PR	24-SEP-1998; 98GB-0020816.
PR	23-JAN-1998; 98US-0082324.
PR	11-JUN-1998; 98GB-0012660.
PR	11-JUN-1998; 98US-0096347.
XX	11-JUN-1998; 98US-0096731.
PA	(COLD-) COLD SPRING HARBOR LAB.
PA	(DEVG-) DEGEN NV.
PX	Bogaert TAOE, Hengartner MO, Liu Q, Van Criekinge WMR;
PX	WPI: 1999-469130/39.
DR	N-PsDB; X99301.
XX	
PT	New isolated Caenorhabditis elegans and human CED-6 genes, used to
PT	develop products for treating, e.g. inflammation or autoimmune
PT	disease
PS	Claim 1a; Fig 2A; 145pp; English.
XX	
CC	The invention provides isolated Caenorhabditis elegans and human CED-6
CC	phagecytosis genes. The nucleic acid sequences encode C. elegans CED-6
CC	polypeptides and human homologues h1CED-6, h2CED-6 and h3CED-6. The CED-6
CC	polypeptides act as adaptor molecules in a signal transduction pathway
CC	that mediates engulfment of apoptotic cells. Transgenic mammalian cells
CC	transfected with the nucleic acids encoding the CED-6 polypeptides can be
CC	used for determining whether a compound is an inhibitor or an enhancer of
CC	a signal transduction pathway which promotes phagocytosis of apoptotic
CC	cells. The enhancers can be used for treating inflammation, autoimmune
CC	disease or cancer. The inhibitors can be used for treating sickle cell
CC	anemia, stroke or neurodegenerative disease. In particular the products
CC	can be used for treating amyotrophic lateral sclerosis, Huntington's
CC	disease, Alzheimer's disease, myocardial heart failure and AIDS. The
CC	products can also be used for detection and diagnosis. The present
CC	sequence represents the amino acid sequence of C. elegans CED-6 protein.
XX	
SQ	Sequence 492 AA;
Query Match	15.0%; Score 98; DB 20; Length 492;
Best Local Similarity	30.8%; Pred. No. 0.14;
Matches 37/ Conservative	15; Mismatches 56; Indels 12; Gaps 5

QY 2 NSKGQVPT--QPIYP-VQPPGNPVYQOTLHLQAPRYTDAPVASELYPSFVHPC--AA 56  
 Db 233 nsKAdyenttppilypgipdplsp---mpqpppnippsisysmprandlptemap 288  
 QY 57 TVPMASAPFGASLYLPMASQSAVAGPLGS--TIPMAVYVPGIYPGSTVLVEGCTDAG 113  
 Db 289 tlpqistsnsqaspsvspastspsgapsippprppalappppvadrinpnvpskntag 348  
 RESULT 11  
 ID P82970  
 P82970 standard; protein; 203 AA.  
 AC P82970;  
 DT 03-DEC-1990 (first entry)  
 DE Bioadhesive precursor protein from CDNA clone 14-1.  
 KW Bioadhesive precursor protein; CDNA clone 14-1; mussels; barnacles;  
 KW oysters; biomedical adhesive; sealants; wound healing.  
 OS Mytilus edulis.  
 PN W08803953-A.  
 PD 02-JUN-1988.  
 PF 24-NOV-1987; 87WO-US03048.  
 PR 07-AUG-1987; 87US-0082456.  
 PA (GENE-) GENEX CORP.  
 PI Maugh KJ, Anderson DM, Strausberg R, Strausberg SL, McCandliss R;  
 DR WPI: 1988-161622/23.  
 DR N-PSDB: N80228.  
 PT DNA coding for bio-adhesive precursor protein - obtd. from mussels,  
 PT barnacles and oysters and used esp. to prepare biomedical adhesive.  
 PS Disclosure; ; pp; English.  
 CC mRNA from the phenol gland of the foot of M. edulis is isolated for  
 CC the synthesis of cDNA. The cDNA is inserted into a bacteriophage  
 CC and E. coli transfectant and cultured. Five clones were isolated and  
 CC characterised, one of which is clone 14-1. The sequence has been  
 CC divided to show 20 subunits, including 19 decapeptides and 1 hexapeptide.  
 CC Clone 14-1 and 52 (N80229) display extensive homology at their 3' ends.  
 CC In particular, the last 138 codons of the translated regions are  
 CC identical and include codons for a hexapeptide followed by 12  
 CC decapeptides (beginning at codon 75 of clone 14-1 and codon 205  
 CC of clone 52. The DNA segment contg. the coding regions of  
 CC lambda gt10 cloning vectors by EcoRI digestion and subcloned into the  
 CC EcoRI site of plasmid pGX2627.  
 CC The protein is chemically or enzymatically hydroxylated to produce  
 CC adhesive compns. having excellent properties in wet environments. The  
 CC product is partic. useful as a biomedical adhesive or sealant e.g. in  
 CC wound healing or as a marine or dental adhesive. It can be used as a  
 CC thin film membrane.  
 CC See also N80228-34.  
 XX  
 XX  
 SQ Sequence 203 AA;  
 Query Match 14.8%; Score 96.5; DB 9; Length 203;  
 Best Local Similarity 27.0%; Pred. No. 0.071;  
 Matches 31; Conservative 8; Mismatches 55; Indels 21; Gaps 4.  
 QY 7 YPTQPTVIVQPPGNPVYQOTLHL-QAPRYTDAPVASELYPSFVHPCAAVPTMSAAF 65

Db 31 ykapsypatcykakpsypptykakpsypptykakpsypptykakpsypptykak 90  
Oy 66 PG-----ASLYLPMAQSVAVGPG-----STIPMAV-----YVGPPIYP 100  
Db 91 psypptykpkpsypptykkskalsypsypkpkklyppptykpkklyppptykpkpsyp 145

## RESULT 12

83193  
ID P83193 standard; Protein; 203 AA.

XX P83193;

DT 06-MAR-1992 (first entry)

DE Sequence of a bioadhesive precursor protein encoded by cDNA clone 14-1.

XX Bioadhesive; wound healing; bonding; recombinant adhesive.

OS Mytilus edulis.

PN W08807076-A.

PD 22-SEP-1988.

PF 11-MAR-1988; 88WO-US00876.

PR 12-MAR-1987; 87US-0025243.

PA (GENE-) GENEX CORP.

PI Maugh KJ, Anderson DM, Strausberg SL, Strausberg R;

PI Wei T;

DR WPI; 1988-285533/40.

DR N-PSDB; N82449.

PT Bio-adhesive precursor protein analogues prodn. - by DNA  
PT recombinant techniques, subsequent hydroxylation, gives an  
PT adhesive for use in wet environment

XX Example; Fig 13; 101pp; English.

XX The bioadhesive precursor protein analogue of the invention can be  
CC hydroxylated (prod. claimed) and used as an adhesive in wet  
CC environments. The resulting bioadhesives have marine, biomedical and  
CC dental applications. They may be used e.g. in wound healing in the  
CC same manner as fibrin. While the decapeptide ala-lys-pro-ser-tyr-pro-  
CC pro-thr-tyr-lys is repeated many times in the polyphenolic adhesive  
CC protein of M.edulis, the examination of cDNA sequences encoding  
CC portions of this protein (see U.S. Patent Application Serial No. 933,  
CC 945) reveals that many other repetitive decapeptide and hexapeptide  
CC sequences are also present in the protein, and that these other  
CC sequences may constitute the majority of AA sequence of the  
CC polyphenolic adhesive protein of M.edulis. For example, in cDNA  
CC clone 14-1 nineteen decapeptides and one hexapeptide are encoded.  
CC  
SQ Sequence 203 AA:

Query Match 14.8%; Score 96.5; DB 9; Length 203;  
Best Local Similarity 27.0%; Pred. No. 0.071;

Matches 31; Conservative 8; Mismatches 55; Indels 21; Gaps 4;

Oy 7 YFTQPTTPYQPGNIPYPTQTLHL-PQAPRYTDAPPAVSSELYRPSFVHPSAAVPTMSAA 65  
Db 31 ykapsypatcykakpsypptykakpsypptykakpsypptykakpsypptykak 90  
Oy 66 PG-----ASLYLPMAQSVAVGPG-----STIPMAV-----YVGPPIYP 100  
Db 91 psypptykpkpsypptykkskalsypsypkpkklyppptykpkklyppptykpkpsyp 145

## RESULT 13

W61954  
ID W61954 standard; protein; 258 AA.

XX W61954;

DT 18-SEP-1998 (first entry)

DE Rat galectin amino acid sequence.

XX Mortalin; galectin; diabetes-mediating protein; insulin; DMP;  
XX diabetes; drug screening assay.

OS Rattus sp.

PN W09820124-A2.

PD 14-MAY-1998.

PF 24-OCT-1997; 97WO-IB01627.

PR 18-JUL-1997; 97US-0897098.

PR 25-OCT-1996; 96US-0029324.

PR 05-NOV-1996; 96US-0030088.

PR 05-NOV-1996; 96US-0030186.

PA (ANDE/) ANDERSEN H U.

PA (CHRI/) BJERRE CHRISTENSEN U.

PA (FEYS/) FEY S J.

PA (KARL/) KARLSEN A E.

PA (LARS/) MOSE LARSEN P.

PA (NERU/) NERUP J.

PA (POCI/) POCIOR F.

XX Andersen HU, Bjerre Christensen U, Fey SJ, Karlsten AE;

XX Mose Larsen P, Nerup J, Pociot F;

XX WPI; 1998-286940/25.

PT Identification of diabetes-mediating protein(s) - by transplanting  
PT insulin-secreting cells into host at risk of developing diabetes and  
PT analysing protein expression in transplanted cells  
XX Disclosure; Fig 4; 154pp; English.

XX This represents the amino acid sequence of murine mortalin. This is a  
XX diabetes-mediating protective protein used in the method of invention.  
XX The invention provides methods for in vivo identification of a diabetes-  
XX mediating protein (DMP) by transplanting insulin-secreting cells into  
XX host at risk of developing diabetes and analysing protein expression in  
XX transplanted cells. The DMPs are useful in drug screening assays for  
XX identifying compounds capable of modulating the development of diabetes,  
XX useful as therapeutic agents for the treatment or prevention of diabetes,  
XX ameliorating diabetes by modulating the expression of the DMP. Changes in  
XX the expression of specific DMPs is diagnostically useful as indicator of  
XX the development of diabetes.

SQ Sequence 258 AA:

Query Match 14.8%; Score 96.5; DB 19; Length 258;  
Best Local Similarity 33.6%; Pred. No. 0.093;

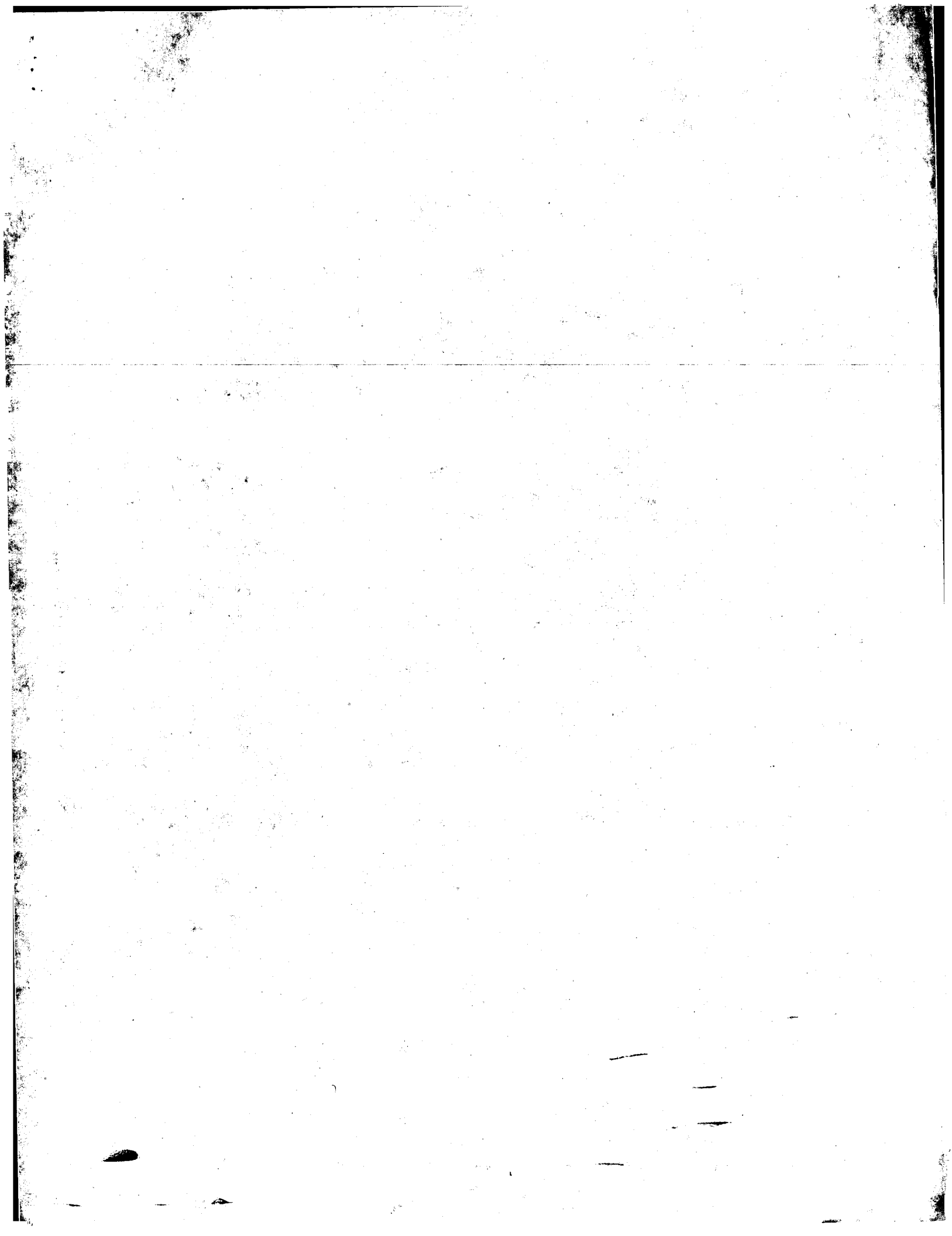
Matches 40; Conservative 2; Mismatches 40; Indels 37; Gaps 8;

Oy 5 GQYP--TQPTYPVQ--PQGNIPYPTQTLHL-PQAPRYTDAPPAVSSELYRPSFVHPSAAVPTMSAA 61  
Db 33 gypypasypgpgpgpgpgpgp--yp-----gqpppsypgpgpgpsay-pgptapggypgpcva 84  
Oy 62 SNAFPGASLYLPMAQSVAVGPGSTIPMAVYPVGP-IYP-PGSTVYLVGQYDAGARFGA 118  
Db 33 gypypasypgpgpgpgpgpgp--yp-----gqpppsypgpgpgpsay-pgptapggypgpcva 84



Tue Mar 6 12:02:32 2001

us-09-407-430-1\_copy\_1\_120.rag





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:52:11 ; Search time 34.57 Seconds  
(Without alignments)  
62.333 Million cell updates/sec

Title: US-09-407-430-1\_COPY\_1\_120  
Perfect score: 653  
Sequence: 1 MNSKGYPTQPTVQVQPGN.....GSTVLVEGGYDAGARFGAGA 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/1aa/6\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/1aa/PCRTUS\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/1aa/backfileall.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	98.5	15.1	5202236-13	Patent No. 5202236
2	98.5	15.1	5202236-25	Patent No. 5202236
3	98	15.0	US-08-562-311-4	Sequence 4, Appl1
4	96.5	14.8	202 5202236-36	Patent No. 5202236
5	96.5	14.8	203 5202236-1	Patent No. 5202236
6	96.5	14.8	262 3 US-08-946-914-14	Sequence 14, Appl1
7	94	14.4	259 3 US-08-476-509B-29	Sequence 29, Appl1
8	94	14.4	574 3 US-09-276-400-6	Sequence 6, Appl1
9	94	14.4	902 1 US-08-396-479B-6	Sequence 6, Appl1
10	94	14.4	902 1 US-08-818-823-6	Sequence 6, Appl1
11	93.5	14.3	331 5 5202236-37	Patent No. 5202236
12	93.5	14.3	334 5 5202236-3	Patent No. 5202236
13	93	14.2	214 1 US-08-217-327-4	Sequence 4, Appl1
14	91.5	14.0	267 1 US-07-706-872-3	Sequence 3, Appl1
15	90.5	13.9	503 3 US-08-526-136-2	Sequence 2, Appl1
16	90.5	13.9	505 3 US-08-526-136-4	Sequence 4, Appl1
17	89	13.6	250 1 US-08-562-311-2	Sequence 2, Appl1
18	89	13.6	250 3 US-08-946-914-10	Sequence 10, Appl1
19	88	13.5	466 3 US-08-526-136-13	Sequence 13, Appl1
20	87.5	13.4	129 5 5202236-5	Patent No. 5202236
21	87	13.3	2476 2 US-08-276-967-2	Sequence 2, Appl1
22	85.5	13.1	2441 1 US-08-194-468-2	Sequence 2, Appl1
23	85.5	13.1	2441 1 US-08-961-739-2	Sequence 2, Appl1
24	84	12.9	509 3 US-08-860-635A-21	Sequence 2, Appl1
25	82.5	12.6	267 3 US-08-301-162-16	Sequence 16, Appl1
26	82.5	12.6	546 1 US-08-494-168-10	Sequence 10, Appl1
27	82	12.6	507 3 US-08-860-635A-19	Sequence 19, Appl1
28	80	12.3	312 3 US-08-894-017-10	Sequence 10, Appl1

29	80	12.3	454	2	US-09-014-969-9	Sequence 9, Appl1
30	80	12.3	1361	3	US-08-894-017-23	Sequence 23, Appl1
31	80	12.3	1565	5	5352450-2	Patent No. 5352450
32	79	12.1	297	2	US-08-580-545B-6	Sequence 6, Appl1
33	79	12.1	297	3	US-09-262-653A-6	Sequence 6, Appl1
34	78.5	12.0	86	5	5202236-7	Patent No. 5202236
35	78.5	12.0	294	3	US-08-742-185-97	Sequence 97, Appl1
36	78	11.9	514	3	US-08-796-899-25	Sequence 25, Appl1
37	78	11.9	1162	2	US-08-728-323A-2	Sequence 2, Appl1
38	77.5	11.9	590	2	US-08-785-310A-5	Sequence 5, Appl1
39	77	11.8	406	1	US-07-973-431B-1	Sequence 1, Appl1
40	77	11.8	1248	2	US-09-080-897-2	Sequence 2, Appl1
41	76.5	11.7	521	4	PCT-US83-08386-10	Sequence 10, Appl1
42	76.5	11.7	539	3	US-08-906-360-1	Sequence 1, Appl1
43	75.5	11.6	380	1	US-08-585-758A-4	Sequence 4, Appl1
44	75.5	11.6	380	1	US-08-977-818-4	Sequence 4, Appl1
45	75.5	11.6	380	2	US-08-670-274B-4	Sequence 4, Appl1

## ALIGNMENTS

```
RESULT 1
5202236-13
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; SEQ ID NO:13:
; LENGTH: 652
5202236-13

Query Match 15.1% Score 98.5; DB 5; Length 652;
Best Local Similarity 28.4%; Pred. No. 0.055;
Matches 31; Conservative 11; Mismatches 56; Indels 11; Gaps 4;

OY 3 SKGYPTQPTVQVQPGNVPQTLHL-PQAPPTDAPPAVSELY--RPSF-----VHPG 54
DB 301 AKPTTKAKSYPTTKAKKSYPTTKAKKSYPTTKAKKSYPTTKAKKSYPTTKAKK 360
OY 55 AATPTMSAIFPGASLYLFMAQSVAVGPIGSIIPAY--YVGPYIP 100
DB 361 YKAKPTPTTKAKKSYPTTKAKKSYPTTKAKKSYPTTKAKKSYPTTKAKKSYPT 409

RESULT 2
5202236-25
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILIPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
```

5 GQYPTQP --- TYPVQ - PPGNPNVYRQILHLRQAPRYTDAPRAYSELYRPSFVHFGAATVPT 60



```

; SEQ ID NO 6
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-276-400-6

```

Query Match	14.4%;	Score	94;	DB	3;	Length	574;
Best Local Similarity	32.7%;	Pred. No.	0.13;				
Matches	34;	Conservative	8;	Mismatches	44;	Indels	18;
						Gaps	6

QY 7 YRPFPTYPQP RPNVVPQT LHL PQA PPTDAPAYSE ---LYRRS -FVHPGAATPYTM 61  
| | : | : | | : | : | : | : | :  
Db 167 YRQPSYPQQP ----YPQABFVSQYENYDPASRTYYDOGFVYIRPAGGCVGA GAALAVASA 222

```

OY      62 SAAFGASLIYPMASVANG---PLGSTIPAIYV--GPIYPP 100
      : | | | | | | | | | | | | | | | | | |
Db 223 GVYIP---YQPRAYEYEGGGEELPEYPPGFGYPAERPYVP 262

```

RESULT 9  
US-08-396-479B-6  
; Sequence 6, Application US/08396479B  
; Patent No. 5612455  
CURRENT INFORMATION:

APPLICANT: HOEY, Timothy  
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEH, HOBACH, TEST, ALBRITTON & HEBBERICH  
STREET: 4 Embarrasadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patent Release #1.0, Version #1.30
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/396,479B

```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

TELEPHONE: (415) 494-8771  
 TELEFAX: (415) 494-8771  
 TELNET: 210 277259  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 902 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-396-479B-6

Query Match	14.4%;	Score 94;	DB 1;	Length 902;
Best Local Similarity	25.3%;	Pred. NO. 0.22;		
Matches 39;	Conservative 11;	Mismatches 56;	Indels 48;	Gaps 6;

QY 8 PTQTPTVPVQPPGNGPVVPTQTLHLHQAAPPYTADAPAYSELKRSFVNHGAAIVP----TMSA 63  
| : | | | : | : | : | | | : | | | : |  
Db 717 PPRPPPPSYPHEDPAC-ETPYLSEGFY-GMPRLYPQTGPPPPSYRGLRMFPDET RGTGGC 714

```

QY      64 AEPGASLIYLPMA-QSVAVGPLGSTITPM----- 89
      | | | : || | | | | | | |
Db      775 AQPAAVSFLRRFPSPDPYGGRGSSFPGLGTFSPAPFRPPPLPASPPLEGGPFPSSQSDVHP 834

```

```

Qy 90 ----AYYPVGPIPPGCVLVE----GGYDAGAR 115
      |  |  |  |  |  |  |  |  |  |  |  |  |
Db 835 LPAEGYNKVGPGYGPGEAPEQEKSRGGYSSGFR 868

```

RESULT 10  
US-08-818-823-6  
; Sequence 6, Application US/08818823  
; Patent No. 5708158

1 APPLICANT: HOEY, Timothy  
 2 TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY  
 3  
 4 NUMBER OF SEQUENCES: 18  
 5  
 6 CORRESPONDENCE ADDRESS:  
 7  
 8 ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HENBERGER  
 9  
 10 STREET: 4 Embarcadero Center, Suite 3400  
 11  
 12 CITY: San Francisco  
 13  
 14 STATE: California  
 15  
 16 COUNTRY: USA  
 17  
 18 ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,823  
FILING DATE: 14-MAR-1997

CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/396,479  
FILING DATE: 02-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59450-1/RACQ  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 210 277299.  
INFORMATION FOR SEQ ID NO: 6  
SEQUENCE CHARACTERISTICS:  
LENGTH: 902 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-818-823-6

Query Match	14.48;	Score 94;	DB 1;	length 902;
Best Local Similarity	25.38;	Pred No. 0.22;		
Matches 39;	Conservative	11;	Mismatches 56;	Indels 48;
			Gaps	6

QY 8 PTQPTPVQPFGNPNVYPQTLLHLPAPRPYTDAAPRAISELYRSEFVNHGATVP----TMSA 63  
| : | | : : : | : | : |  
Db 717 PRPRPVSYPNHEDAC-ETPYLSEGFGY-GMPRLYPQTGPPPSYRGLRMFEFTRGTTCG 774

```

QY      64  AFGASLILPMA-QSVAVGPLGSITPM-----89
          | | : || | | | :
Db      775  AQPPAVSFLRPPPSDPYGGRSSPLGLDFSPPADFRPPLPASPPLEGPPSSQSDVHP83

```

```

Oy      90 ----AYYPVGPIYPGGSTLVE----GGYDAGAR 115
          |  |||  |  |  :  |||  :  |  |
Db      835 LPAEGYNKVGPGYGPGECAPEGEKSRGGYSSGFR 866

```

RESULT	11
5202236-37	
Patent NO	5202236

/ FILING NO.: 20220805  
 / APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG  
 / SUSAN L.; MCCANDLISS, RUSS; WEL, TENA; FILIPULA, DAVID  
 / TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
 / PROTEIN

RESULT 11  
5202236-37  
; Patent NO. 5202236  
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
; SUSAN L.; MCCANDLISH, RUSS; WEL, TENA; FILIPULA, DAVID  
; TITLE OF INVENTION: METHOD OF PRODUCING BIOMIMETIC  
; PROTEIN

NUMBER OF SEQUENCES: 39  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/528,762  
FILING DATE: 25-MAY-1990  
APPLICATION NUMBER: 82,456  
FILING DATE: 07-AUG-1987  
APPLICATION NUMBER: 933,945  
FILING DATE: 24-NOV-1986  
APPLICATION NUMBER: 650,128  
FILING DATE: 13-SEP-1984  
SEQ ID NO: 37  
LENGTH: 331  
5202236-37

Query Match 14.3% Score 93.5; DB 5; Length 331;  
Best Local Similarity 30.8%; Pred. No. 0.075;  
Matches 33; Conservative 6; Mismatches 51; Indels 17; Gaps 5;

QY 3 SKGOYPTQPTTYVQPPG--NPVYPTQLHL-----PQAPPYTDAPPVSELYRPSFVHP 53  
DB 175 AKRSYP--PTYKAKPTYYKIKPTPTSTYKAKPTYYKAKPTYYKAKPSYPTYYKPKPSYP 232  
QY 54 GAATVPTMSAARPGASLYLPMASVAVGCLGSTITPMATYVPGIYYP 100  
DB 233 -----PTYKSKSTYPSYKPKTKTPTYYKPKLTYPPTYKRP--KPSYYP 273

RESULT 12  
5202236-3  
PATENT NO. 5202236  
APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,  
SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID  
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE  
PROTEIN  
NUMBER OF SEQUENCES: 39  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/528,762  
FILING DATE: 25-MAY-1990  
APPLICATION NUMBER: 82,456  
FILING DATE: 07-AUG-1987  
APPLICATION NUMBER: 933,945  
FILING DATE: 24-NOV-1986  
APPLICATION NUMBER: 650,128  
FILING DATE: 13-SEP-1984  
SEQ ID NO: 3  
LENGTH: 334  
5202236-3

Query Match 14.3% Score 93.5; DB 5; Length 334;  
Best Local Similarity 30.8%; Pred. No. 0.076;  
Matches 33; Conservative 6; Mismatches 51; Indels 17; Gaps 5;  
QY 3 SKGOYPTQPTTYVQPPG--NPVYPTQLHL-----PQAPPYTDAPPVSELYRPSFVHP 53  
DB 178 AKRSYP--PTYKAKPTYYKIKPTPTSTYKAKPTYYKAKPTYYKAKPSYPTYYKPKPSYP 235  
QY 54 GAATVPTMSAARPGASLYLPMASVAVGCLGSTITPMATYVPGIYYP 100  
DB 236 -----PTYKSKSTYPSYKPKTKTPTYYKPKLTYPPTYKRP--KPSYYP 276

RESULT 13  
US-08-217-327-4  
Sequence 4, Application US/08217327  
PATENT NO. 5474925  
GENERAL INFORMATION:  
APPLICANT: John, Malyakal E  
TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles and Brady  
STREET: P.O. Box 2113  
CITY: Madison  
STATE: WI  
COUNTRY: USA  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/217,327  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/812,233  
FILING DATE: 19-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 1122990831  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 214 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-217-327-4

Query Match 14.2% Score 93; DB 1; Length 214;  
Best Local Similarity 29.6%; Pred. No. 0.051;  
Matches 29; Conservative 10; Mismatches 51; Indels 8; Gaps 3;  
QY 5 GQPTQPTTYVQPPG--NPVYPTQLHL-----PQAPPYTDAPPVSELYRPSFVHPGAATVPTMSA 64  
DB 23 GOAPSNP--PTSTPATPPPPAS---TPEPTTQARPTPTATPPPVSTPTPTTSSPPPVYAS 76  
QY 65 PGASLYLPMASVAVGCLGSTITPMATYVPGIYYPGS 102  
DB 77 PPPVSTPPSPPPATPPASPPATPPASP--PPAT 112

RESULT 14  
US-07-706-872-3  
Sequence 3, Application US/07706872  
PATENT NO. 5237056  
GENERAL INFORMATION:  
APPLICANT: Fischbach, Gerald D.  
TITLE OF INVENTION: Identification of a Protein Which  
PROMOTES THE SYNTHESIS OF ACETYLCHOLINE RECEPTORS AND USCS  
TITLE OF INVENTION: Therefore  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/706,872  
FILING DATE: 19910529  
CLASSIFICATION: 514

```

1      ATTORNEY/AGENT INFORMATION:
2      NAME: Granahan, Patricia
3      REGISTRATION NUMBER: 32,227
4      REFERENCE/DOCKET NUMBER: H991-01
5      TELECOMMUNICATION INFORMATION:
6      TELEPHONE: (617) 861-6240
7      TELEFAX: (617) 861-9540
8      INFORMATION FOR SEQ ID NO: 3:
9      SEQUENCE CHARACTERISTICS:
10     LENGTH: 267 amino acids
11     TYPE: AMINO ACID
12     TOPOLOGY: linear
13     MOLECULE TYPE: protein
14
15  --
16  US-07-706-872-3

```

Query Match	14.0%;	Score 91.5;	DB 1;	length 267;
Best Local Similarity	24.7%;	Pred. No. 0.092;		
Matches	38;	Conservative	7;	Mismatches 44; Indels 65; Gaps 7.

[illegible]

RESULT 15  
US-08-526-136-2  
: Sequence 2, Application US/08526136

GENERAL INFORMATION:  
APPLICANT: Towle, Christine A. et al  
TITLE OF INVENTION: ANNEXIN XI  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:

ADDRESS: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804

```

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 50Z or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,136

```

CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/214, 036  
 FILING DATE:  
 APPLICATION NUMBER: 07/837, 775  
 FILING DATE: February 13, 1992  
 APPLICATION NUMBER: 07/764, 465  
 FILING DATE: September 23, 1991  
 ATTORNEY/AGENT INFORMATION:  
 \_\_\_\_\_

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 503
; TYPE: amino acid
; TOPOLOGY: N/A
;
US-08-526-136-2

```

Query Match	13.9%;	Score 90.5;	DB 3;	length 503;
Best Local Similarity	29.5%;	Pred. No. 0.24;		
Matches	38;	Conservative	10;	Mismatches 40;
				Indels 41;
				Gaps 9;

QY 5 GQYLPQ-----PRYVQPP--GNPVYPTPLPLQAAPRYTDA-----PAVSLAYSPFVNHG 54

Db 90 GQPPPAQOPVPSYGVYPPPGGNP-----TSGMSPSTPLPGAPVPCQPMPLPPGQQPCGVYPC 145

QY 55 AATVTSMAAFGASLTYPMQSVAVGPIGSLTIMAVY-----VGPIYPPG-----S 102

Db 146 Q---PPMT---YFGQSPVPPPGQ-----PVSYPGYSGSGVTLPANVSQAQFQNRG 190

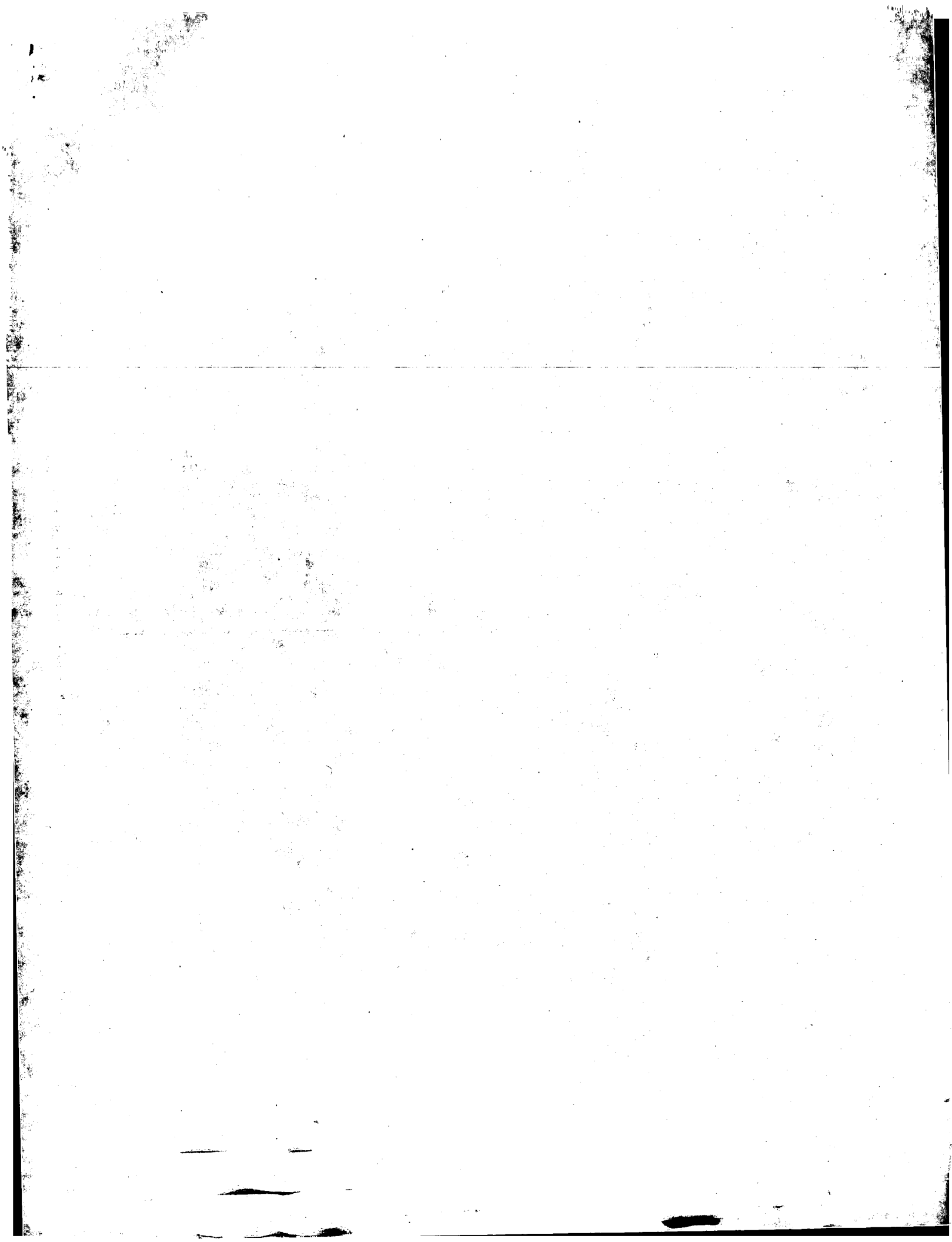
QY 103 TVLVEGGYD 111

Db 191 TTIDAGSFD 199

```
Search completed: March 6, 2001, 11:52:12
Job time: 90 sec
```

Tue Mar 6 12:02:32 2001

us-09-407-430-1\_copy\_1\_120.rai





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:02 ; Search time 45.68 seconds

(without alignments)  
178.373 Million cell updates/sec

Title: US-09-407-430-1\_COPY\_1\_120

Perfect score: 653

Sequence: 1 MNSKQGYPTQPTVPVQPCGN.....GSTVLVEGGYDAGARGACA 120

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891.seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	18.5	491	SI4182	DNA-directed RNA p
2	120.5	18.5	625	T07796	DNA-directed RNA p
3	120.5	18.5	650	SI4181	DNA-directed RNA p
4	120.5	18.5	977	SI4183	DNA-directed RNA p
5	118	18.1	1834	JDMU1	DNA-directed RNA p
6	117.5	18.0	1896	1	DNA-directed RNA p
7	116.5	17.8	1553	TJ1436	DNA-directed RNA p
8	116.5	17.8	1553	1	DNA-directed RNA p
9	115.5	17.7	1690	1	DNA-directed RNA p
10	113	17.3	1605	TJ1435	DNA-directed RNA p
11	111.5	17.1	237	2	DNA-directed RNA p
12	111.5	17.1	467	A27677	DNA-directed RNA p
13	111.5	17.1	1932	1	DNA-directed RNA p
14	111.5	17.1	1970	1	DNA-directed RNA p
15	111.5	17.1	1970	1	DNA-directed RNA p
16	111	17.0	1859	1	DNA-directed RNA p
17	110.5	16.9	1862	2	DNA-directed RNA p
18	110.5	16.9	502	2	DNA-directed RNA p
19	107.5	16.5	620	1	DNA-directed RNA p
20	107.5	16.5	708	1	DNA-directed RNA p
21	107.5	16.5	856	2	DNA-directed RNA p
22	107.5	16.5	902	2	DNA-directed RNA p
23	107	16.4	1752	1	DNA-directed RNA p
24	105.5	15.9	279	2	DNA-directed RNA p
25	104	15.2	346	2	DNA-directed RNA p
26	103.5	15.8	2457	2	DNA-directed RNA p
27	103	15.7	228	2	DNA-directed RNA p
28	102.5	15.7	530	2	DNA-directed RNA p
29	101	15.5	990	2	DNA-directed RNA p

30	101	15.5	2452	1	RN202L	DNA-directed RNA p
31	100.5	15.4	428	2	T24769	hypothetical prote
32	100	15.3	268	2	A49203	homeotic protein C
33	99.5	15.2	32	2	B48831	vitelline membrane
34	99.5	15.2	275	2	T51437	hypothetical prote
35	99.5	15.2	283	2	S13383	hydroxyproline-ric
36	99.5	15.2	751	2	S68957	adhesive plaque pr
37	98	15.0	264	2	A45983	lactose-binding pr
38	98	15.0	481	2	T16484	hypothetical prote
39	98	15.0	492	2	T43064	hypothetical prote
40	97.5	14.9	297	2	S23737	hypothetical prote
41	97	14.9	419	2	T29266	hypothetical prote
42	96.5	14.8	262	2	A54889	hypothetical prote
43	96.5	14.8	532	2	T30764	hypothetical prote
44	96.5	14.8	875	2	S23760	hypothetical prote
45	96	14.7	629	2	T05425	polyadenylate-bind

## ALIGNMENTS

## RESULT

SI4182

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain (isoform B2) - soybean (fragme

C:Species: Glycine max (soybean)

C:Date: 12-Feb-1993 #sequence-revision 12-Feb-1993 #text-change 12-Sep-1997

C:Accession: SI4182

R:Dieterich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A

A:Reference number: SI1960; M01D:91355869

A:Accession: SI4182

A:Molecule type: mRNA

A:Residues: 1-491 <DIE>

A:Cross-references: EMBL:X52493

A>Note: translation of the nucleotide sequence is not complete

A:Gene: RPB1-B2

A:Intons: 416/2; 440/2; 458/2

C:Superfamily: human DNA-directed RNA polymerase II largest chain

C:Keywords: DNA binding; nucleic acidyltransferase; tandem repeat; transcriptio; zinc f

F:201-474/Region: 7-residue repeats

Query Match

Best Local Similarity 32.7%; Score 121; DB 2; Length 491;

Matches 35; Conservative 11; Mismatches 53; Indels 8; Gaps 4;

Qy 3 SKGQYPTQPTVPVQPG---NPVYPT--LHLPAQPPYTDAPPVSELYRPSFVHGAA 56

Db 239 SPQSP 297

Qy 57 TWPVMSAATPGASLYLPMASVA--VGPLGRTTPMAYPVGVGPPGS 102

Db 298 YSP 344

RESULT 2

T07796

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain - soybean (fragment)

C:Species: Glycine max (soybean)

C:Date: 14-May-1999 #sequence-revision 14-May-1999 #text-change 08-Oct-1999

C:Accession: T07796

R:Dieterich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A

A:Reference number: SI1960; M01D:91355869

A:Accession: T07796

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-625 <DIE>

A:Cross-references: EMBL:X52492; NID:g18731; PIDN:CA3673.1; PID:g18732

A:Experimental source: strain var. Wayne









Tue Mar 6 12:02:32 2001

us-09-407-430-1\_copy\_1\_120.rpt

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:31 ; Search time 25.34 Seconds  
(Without alignments)  
152.932 Million cell updates/sec

Title: US-09-407-430-1\_COPY\_1\_120  
Perfect score: 653  
Sequence: 1 MNSKQVPTQPTYPVQPPGN.....GSTVLVEGGYDAGARFGAGA 120

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues  
Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	18.1	1841	RPB1_ARATH	P18616 arabidopsis
2	118	18.1	1860	RPB0_ARATH	P31635 arabidopsis
3	117.5	18.0	1896	RPB1_DROME	P04052 drosophila
4	117	17.9	467	CBPA_DICDI	P35085 dictyostell
5	116.5	17.8	1733	RPB1_YEAST	P04050 saccharomyc
6	111.5	17.1	467	RPB1_CRIGR	P14144 cricetulus
7	111.5	17.1	1970	RPB1_HUMAN	P24328 homo sapien
8	111.5	17.1	1970	RPB1_HUMAN	P08775 mus musculu
9	111	17.0	1859	RPB1_CAEEL	P16356 caenorhabdi
10	107.5	16.5	620	EXTN_TOBAC	P13983 nicotiana t
11	107.5	16.5	708	VP40_HCMVA	P16753 human cytom
12	107.5	16.5	902	RPB1_DICDI	P35084 dictyostell
13	107	16.4	1752	RPB1_SCHPO	P36594 schizosacch
14	104	15.9	346	PRE1_LCES	Q00451 lycopersico
15	101	15.3	2452	RPB1_PLAFD	P14248 plasmodium
16	100.5	15.4	872	FPL_MYTGO	Q25434 mytilus cor
17	100	15.3	268	CDX1_MOUSE	P18111 mus musculu
18	99.5	15.2	283	EXTN_SORBI	P24152 sorghum bic
19	99.5	15.2	751	FPL_MYTGA	Q27409 mytilus gal
20	96.5	14.8	261	LEG3_RAT	P08699 ratulus norv
21	96.5	14.8	875	FPL_MYTED	Q25460 ratulus norv
22	96	14.7	629	PAB2_ARATH	P42731 arabidopsis
23	94.5	14.5	902	IF2_MYTCTU	P71613 mycobacteri
24	94	14.4	902	NFCA_HUMAN	Q14934 homo sapien
25	93	14.2	295	LEG3_CANFA	P34846 canis famli
26	93	14.2	503	ANYA_RABIT	P33477 oryctolagus
27	92	14.1	541	TELL_DROME	P09957 drosophila
28	91	13.9	503	ANYA_MOUSE	P97384 mus musculu
29	91	13.9	1048	SR44_RAT	Q63627 ratulus norv
30	90.5	13.9	503	ANYA_BOVIN	P27214 bos taurus
31	90.5	13.9	505	ANXB_BOVIN	P27215 bos taurus
32	90.5	13.9	1183	DREL_RAT	P52258 ratulus norv
33	90.5	13.9	2142	BAT2_HUMAN	P48634 homo sapien

## ALIGNMENTS

RESULT . 1	ID	RPB1_ARATH	STANDARD:	PRT: 1841 AA.	
AC	P18616;				
DT	01-NOV-1990 (Rel. 16, Created)				
DT	01-NOV-1990 (Rel. 16, Last sequence update)				
DT	15-DEC-1998 (Rel. 37, Last annotation update)				
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)				
DE	(VERSION 1)				
GN	RPB205 OR RPB1 OR RPB1.				
OS	Arabidopsis thaliana (Mouse-ear cress).				P16110 mus musculu
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				P39061 mus musculu
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;				P17931 homo sapien
OC	Brassicales; Brassicaceae; Arabidopsis.				P18488 drosophila
RN	[1]				P03330 simlen sarc
RP	SEQUENCE FROM N.A.				P27177 gallus gall
RC	STRATIN-CV. COLUMBIA;				P17277 gallus gall
RX	MEDLINE-9108067; PubMed-2259344;				P09125 elmeria ace
RA	Nawrath C., Schell J., Koncz C.;				P47902 homo sapien
RT	"Homologous domains of the largest subunit of eucaryotic RNA				P20073 homo sapien
RT	polymerase II are conserved in plants.";				O95104 homo sapien
RL	Mol. Gen. Genet. 223:65-75(1990).				O15178 homo sapien
CC	-1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION				
CC	OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS				
CC	SUBSTRATES.				
CC	-1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +				
CC	RNA(N).				
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.				
CC	-1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.				
CC	THE PHOSPHORYLATION ACTIVATES POL2.				
CC	-1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE				
CC	FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA				
CC	PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE				
CC	III FOR 5S AND TRNA GENES.				
CC	-1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION				
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION				
CC	AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS				
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY				
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL				
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	OR SEND AN EMAIL TO <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
DR	EMBL: X52954; CAA37130.1; -				
DR	PIR: S12071; JDM01.				
DR	INTERPRO: IPR000684; -				
DR	INTERPRO: IPR000722; -				
DR	INTERPRO: IPR002879; -				
DR	PFAM: PF01854; RNA_POL_A2; 1.				
DR	PFAM: PF00623; RNA_POL_A; 1.				
DR	PROSITE: PS00115; RNA_POL_II_REPEAT; 24.				
KW	Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;				
KW	DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.				
FT	ZN_FING 66 79				
FT	DNA_BIND 327 398				
FT	BY SIMILARITY.				





DR EMBL; 003413; AAA03471.1; -  
DR DICTYDB; DD01055; CBPA.

CC RNA(N).  
CC -1 SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
CC -1 SUBCELLULAR LOCATION: NUCLEAR.  
CC -1 PWM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
CC THE PHOSPHORYLATION ACTIVATES POL2.  
CC -1 MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
CC

CC PRECURSOR, POLYMERASE II FOR THE mRNA PRECURSOR, AND POLYMERASE  
CC III FOR 5S AND TRNA GENES.  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X03128; CAA26904.1; -  
CC DR EMBL: X96876; CAA65619.1; -  
CC DR EMBL: Z74188; CAA98713.1; -  
CC DR EMBL: U27182; AAC49058.1; -  
CC DR PIR: A00692; RNEY2L.  
CC SCD: S0002299; PRO21.  
CC INTERPRO: IPR000684; -  
CC INTERPRO: IPR000722; -  
CC DR INTERPRO: IPR002879; -  
CC DR PFAM: PF01854; RNA\_POL\_A2; 1.  
CC DR PFAM: PF00623; RNA\_POL\_A; 1.  
CC DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 22.  
CC KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
CC KM DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
CC FT ZN-FING 67 83  
CC FT DOMAIN 1544 1719 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
CC FT VARIANT 1553 1659 MISSING (IN STRAIN A364A).  
CC FT CONFLICT 1514 1514 A -> V (IN REF. 1).  
CC FT CONFLICT 1524 1524 G -> A (IN REF. 1).  
CC FT CONFLICT 1601 1601 T -> M (IN REF. 1).  
CC SQ SEQUENCE 1733 AA; 191610 MW; A45C1360F99F968 CRC64;  
  
Query Match 17.8%; Score 116.5; DB 1; Length 1733;  
Best Local Similarity 31.7%; Pred. No. 0.092;  
Matches 33; Conservative 14; Mismatches 52; Indels 5; Gaps 4;  
  
OY 1 MNSKGYPTQPTTY-PVQPPGNPVYQTLHLPOAPYTDAPPAYSELYRSPFVHPCATVPP 59  
DB 1537 VSSPGSP 1593  
DB 60 TMSAAPPASLYIPMAQSYA-VGPIGSTIPMAVYPVGPPIYP 102  
DB 1594 TSPSYSP 1637  
  
RESULT 6  
RPBL\_CRIGR STANDARD; PRT; 467 AA.  
ID RPBL\_CRIGR  
AC P11414;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1)  
DE (FRAGMENT).  
GN POLR2A.  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetulus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-88094402; PubMed-3122024;  
RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.,  
RT "The C-terminal domain of the largest subunit of RNA polymerase II of  
RT Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a  
RT conserved structure with an essential function.";  
RL Mol. Cell. Biol. 8:321-329(1988).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
CC RNA(N).  
CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
CC THE PHOSPHORYLATION ACTIVATES POL2.  
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
CC PRECURSOR, POLYMERASE II FOR THE mRNA PRECURSOR, AND POLYMERASE  
CC III FOR 5S AND TRNA GENES.  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M19538; AAA37008.1; -  
CC DR PIR: A27677; A27677.  
CC DR INTERPRO: IPR000684; -  
CC DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 43.  
CC KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
CC KM DNA-binding; Nuclear protein; Phosphorylation.  
CC FT NON\_TER 1 457  
CC FT DOMAIN 90 457 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
CC SQ SEQUENCE 467 AA; 48167 MW; 0AADD923B5D2BCE CRC64;  
  
Query Match 17.1%; Score 111.5; DB 1; Length 467;  
Best Local Similarity 31.6%; Pred. No. 0.062;  
Matches 30; Conservative 12; Mismatches 48; Indels 5; Gaps 4;  
  
OY 8 PTQPTTY-PVQPPGNPVYQTLHLPOAPYTDAPPAYSELYRSPFVHPCATVPPMSAFP 66  
DB 289 PTPSP 345  
DB 346 SSEPYPSP 380  
  
RESULT 7  
RPBL\_HUMAN STANDARD; PRT; 1970 AA.  
ID RPBL\_HUMAN  
AC P24928;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPB1).  
DE POLR2A.  
GN Homo sapiens (Human).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92178992; PubMed-1542581;  
RA Wintzerith M., Acker J., Vicalre S., Vigneron M., Kedinger C.,  
RT "Complete sequence of the human RNA polymerase II largest subunit.";  
RT Nucleic Acids Res. 20:910-910(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-95347616; PubMed-7622068;  
RA Mita K., Tsuji H., Moriyomo M., Takahashi E., Nenoi M.,  
RA Ichimura S., Yamachi M., Hong E., Haysashi A.,  
RT "The human gene encoding the largest subunit of RNA polymerase II.";  
RL Gene 159:285-286(1995).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +

CC RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 CC EMBL: X63564; CAA5125.1; -  
 CC EMBL: X74874; CAA52862.1; -  
 CC EMBL: X74873; CAA52862.1; JOINED.  
 CC EMBL: X74872; CAA52862.1; JOINED.  
 CC EMBL: X74871; CAA52862.1; JOINED.  
 CC EMBL: X74870; CAA52862.1; JOINED.  
 CC PIR: S21054; S21054.  
 CC MIM: 180660; -  
 CC INTERPRO: IPR000684; -  
 CC INTERPRO: IPR000722; -  
 CC INTERPRO: IPR002879; -  
 CC PFM: PF01854; RNA\_POL\_A2; 1.  
 CC PFM: PF00623; RNA\_POL\_A; 1.  
 CC PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 43.  
 CC TRANSFASER: DNA-directed RNA polymerase; Transcription; zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.  
 CC ZN-FING 71 87  
 CC DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC CONFLICT 1067 1067 W -> L (IN REF. 2).  
 CC CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 CC FT CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 CC FT CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 CC SQ SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;

Query Match 17.1%; Score 111.5; DB 1; Length 1970;  
 Best Local Similarity 31.6%; Pred. No. 0.25; Mismatches 48; Indels 5; Gaps 4;  
 Matches 30; Conservative 12; Mismatches 48; Indels 5; Gaps 4;  
 Oy 8 PTOPTV-VVPGPGNPVYPTQTLHLFQADPPYDAPPAVSELYRPSFVHPGATVPTMSAAPP 66  
 Db 1792 PTSPSYSTSPSSPSR-YTPQSPTYTPSSPSYSP-SSPSYSPSPKVTPTSPSYSP 1848  
 Oy 67 GASLYLPMAQSA-VGFLGRTIPMAVYVPGPIYP 100  
 Db 1849 SSPETPTSPKYSPTSPKYSPTSPKYSPTSPSYSP 1883

RESULT 8  
 ID REP1\_MOUSE STANDARD; PRT; 1970 AA.  
 AC P08775;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 15-MAR-1992 (Rel. 21, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (REP1).  
 GN POLR2A OR RPO2-1 OR RPII215.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-87280135; PubMed-3038894;  
 RA Ahern J.M. Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;  
 RT "Cloning and sequence analysis of the mouse genomic locus encoding  
 the largest subunit of RNA polymerase II.";

RL J. Biol. Chem. 262:10695-10705(1987).  
 RN (2)  
 RP SEQUENCE OF 1587-1970 FROM N.A.  
 RX MEDLINE-86068017; PubMed-2999785.  
 RA Corden J.L., Cadena D.L., Ahern J.M. Jr., Dahmus M.E.;  
 RT "A unique structure at the carboxyl terminus of the largest subunit  
 of eukaryotic RNA polymerase II.";  
 Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).  
 RN (3)  
 RP SEQUENCE OF 32-1970 FROM N.A.  
 RA Ahern J.M. Jr.;  
 RL Submitted (SEP-1986) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.  
 RX MEDLINE-92178992; PubMed-1542581.  
 RA Wintzerlich M., Acker J., Vicaire S., Vigneron M., Kedinger C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit."  
 Nucleic Acids Res. 20:910-910(1992).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 CC EMBL: M12130; AAA40071.1; -  
 CC EMBL: M14101; AAA40071.1; JOINED.  
 CC PIR: A28490; A28490.  
 CC MGD: MGI:98086; RPO2-1.  
 CC INTERPRO: IPR000684; -  
 CC INTERPRO: IPR000722; -  
 CC INTERPRO: IPR002879; -  
 CC PFM: PF01854; RNA\_POL\_A2; 1.  
 CC PFM: PF00623; RNA\_POL\_A; 1.  
 CC PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 42.  
 CC TRANSFASER: DNA-directed RNA polymerase; Transcription; zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.  
 CC ZN-FING 71 87  
 CC DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC CONFLICT 1498 1498 P -> R (IN AAA40071).  
 CC FT CONFLICT 1499 1536 MISSING (IN AAA40071).  
 CC SQ SEQUENCE 1970 AA; 217175 MW; 7D76F38FD92A657E CRC64;

Query Match 17.1%; Score 111.5; DB 1; Length 1970;  
 Best Local Similarity 31.6%; Pred. No. 0.25; Mismatches 48; Indels 5; Gaps 4;  
 Matches 30; Conservative 12; Mismatches 48; Indels 5; Gaps 4;  
 Oy 8 PTOPTV-VVPGPGNPVYPTQTLHLFQADPPYDAPPAVSELYRPSFVHPGATVPTMSAAPP 66  
 Db 1792 PTSPSYSTSPSSPSR-YTPQSPTYTPSSPSYSP-SSPSYSPSPKVTPTSPSYSP 1848  
 Oy 67 GASLYLPMAQSA-VGFLGRTIPMAVYVPGPIYP 100  
 Db 1849 SSPETPTSPKYSPTSPKYSPTSPKYSPTSPSYSP 1883

RESULT 9  
 ID REBEL CAEEL STANDARD: PRT: 1859 AA.  
 AC P16356;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6).  
 GN AMA-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE-9006416; PubMed-2586513;  
 RA Bird D.M., Riddle D.L.;  
 RT "Molecular cloning and sequencing of ama-1, the gene encoding the  
 RT largest subunit of Caenorhabditis elegans RNA polymerase II.";  
 RL Mol. Cell. Biol. 9:4119-4130(1989).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
 CC RNA(N).  
 CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM NINE TO  
 CC FOURTEEN DIFFERENT POLYPEPTIDES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND 5.8S GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: M29235; AAA28126.1; -  
 CC DR PIR: A34092; A34092.  
 CC DR INTERPRO: IPR000684; -  
 CC DR INTERPRO: IPR000722; -  
 CC DR INTERPRO: IPR002879; -  
 CC DR PIR: P01854; RNA\_POL\_A2.1.  
 CC DR PIR: P00623; RNA\_POL\_A1.1.  
 CC DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT. 26.  
 CC KW Transferrase; DNA-directed RNA polymerase; transcription; zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.  
 CC FT ZN.FING 66 82 C2H2-TYPE (POTENTIAL).  
 CC FT DOMAIN 1567 1859 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC FT SEQUENCE 1859 AA; 204631 MW; 756F687B0AACCFC8 CRC64;  
 CC SQ  
 CC  
 CC Query Match 17.0%; Score 111; DB 1; Length 1859;  
 CC Best Local Similarity 29.5%; Pred. No. 0.26;  
 CC Matches 33; Conservative 14; Mismatches 57; Indels 8; Gaps 5;

EXTN\_TOBAC  
 ID EXTN\_TOBAC STANDARD: PRT: 620 AA.  
 AC P13983;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-MAR-1992 (Rel. 21, Last annotation update)  
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).  
 GN HRPPT3.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; easterids I;  
 OC Solanales; Solanaceae; Nicotiana.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. XANTHI; TISSUE-LEAF;  
 RX MEDLINE-90128263; PubMed-2612909;  
 RA Keller B., Lamb C.J.;  
 RT "Specific expression of a novel cell wall hydroxyproline-rich  
 RT glycoprotein gene in lateral root initiation.";  
 RL Genes Dev. 3:1639-1646(1989).  
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN  
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE  
 CC MAIN ROOT.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE  
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
 CC GLYCOSYLATED.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: X13885; CAA32090.1; -  
 CC DR PIR: S06733; S06733.  
 CC KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
 CC FT SIGNAL 1 2 3  
 CC FT CHAIN 1 2 3  
 CC FT REPEAT 70 73  
 CC FT REPEAT 148 151 H-A-P-P.  
 CC FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.  
 CC FT REPEAT 229 235 1.  
 CC FT REPEAT 236 242 2.  
 CC FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.  
 CC FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.  
 CC FT SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;  
 CC SQ  
 CC  
 CC Query Match 16.5%; Score 107.5; DB 1; Length 620;  
 CC Best Local Similarity 28.9%; Pred. No. 0.16;  
 CC Matches 33; Conservative 15; Mismatches 49; Indels 17; Gaps 4;

DE		CAPSID ASSEMBLY PROTEIN).
GN	UB80 OR APNG.	
OS	Human cytomegalovirus (strain AD169).	
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;	
CC	Betaherpesvirinae; Cytomegalovirus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	MEDLINE=90260939; PubMed=2613119;	
RX	Chee M.S., Bankier A.T., Beck S., Bohm R., Brown C.M., Cerny R.,	
RA	Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,	
RA	Predick E., Satchwell S.C., Tomlinson P., Weston K.M., Barrill B.G.;	
RT	"Analysis of the protein-coding content of the sequence of human	
RL	cytomegalovirus strain AD169.";	
RN	Curr. Top. Microbiol. Immunol. 154:125-169(1990).	
RM	[2]	
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF PROTEASE.	
RX	MEDLINE=96399135; PubMed=8805706;	
RA	Tong L., Qian C., Massariol M.-J., Bonneau P.R., Cordingley M.G.,	
RA	Lagace L.;	
RT	"A new serine-protease fold revealed by the crystal structure of	
RL	human cytomegalovirus protease.";	
RN	Nature 383:272-275(1996).	
RP	[3]	
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.	
RX	MEDLINE=96399136; PubMed=8805707;	
RA	Qiu X., Calp J.S., Dilalla A.G., Hellmig B., Hoog S.S., Janson C.A.,	
RA	Smith W.W., Abdel-Meguid S.A.;	
RT	"Unique fold and active site in cytomegalovirus protease.";	
RL	Nature 383:275-279(1996).	
RN	[4]	
RP	X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.	
RX	MEDLINE=96399137; PubMed=8805708;	
RA	Shieh H.-S., Kurumbail R.G., Stevens A.N., Stegeman R.A.,	
RA	Sturman E.J., Pak J.Y., Wiltner A.J., Palmer M.O., Wiegand R.C.,	
RA	Holtwerda B.C., Stallings W.C.;	
RT	"Three-dimensional structure of human cytomegalovirus protease.";	
RL	Nature 383:279-282(1996).	
CC	-I- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS A COMPONENT OF THE CAPSID	
CC	CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.	
CC	ASSEMBLIN IS A PROTEASE WHICH CAN PROTOLOGICALLY CLEAVE ITSELF	
CC	AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.	
CC	-I- CATALYTIC ACTIVITY: PREFERENTIALLY CLEAVES AT ALA-I-SER OR ALA-I-	
CC	ALA.	
CC	-I- PRIM: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.	
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21 (SERINE PROTEASE).	
CC	-I- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outpost on	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/isb-sib.ch">http://www.isb-sib.ch/announce/isb-sib.ch</a> ).	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .	
CC	-----	
DR	EMBL: X17403; CA335353.1; -	
DR	EMBL: X17403; CA335354.1; ALT_INIT.	
DR	PIR: S09843; OQBEBB.	
DR	PDB: 1WPO; 15-OCT-97.	
DR	PDB: 1LAY; 26-SEP-97.	
DR	PDB: 1CMV; 04-SEP-97.	
DR	MEROPS: S21.002; -	
DR	INTERPRO: IPR001847; -	
DR	PRIM: PF00716; Peptidase_S21.1.	
DR	PRINTS: PR00236; HSVCAPSIDP40.	
KW	Coat protein; Hydrolase; Serine protease; Phosphorylation;	
FT	3D-structure.	
FT	CHAIN 1 256	ASSEMBLIN (PROTEASE).
FT	CHAIN 257 643	CAPSID ASSEMBLY PROTEIN (BY SIMILARITY).
FT	PROPEP 644 708	C-TERMINAL PEPTIDE (BY SIMILARITY).
FT	CHAIN 1 708	GENE UB80 PROTEIN.
FT	CHAIN 336 708	GENE UB80.5 PROTEIN.
FT	CHAIN 393 708	GENE UB80.4 PROTEIN.

```

FT CHAIN      478    708    GENE U180.3 PROTEIN.
FT SITE       256    257    CLEAVAGE (BY THE PROTEASE).
FT SITE       643    644    CLEAVAGE (BY THE PROTEASE) . (PROBABLE).
FT ACT_SITE   63     63     CHARGE RELAY SYSTEM.
FT ACT_SITE   132    132    CHARGE RELAY SYSTEM.
FT ACT_SITE   157    157    CHARGE RELAY SYSTEM.
SQ SEQUENCE   708 AA: 73851 MW: 32A993D656824C9 CRC64;

Query Match          16.5% Score 107.5; DB 1; Length 708;
Best Local Similarity 28.5%; Pred. No.0.19;
Matches 39; Conservative 14; Mismatches 43; Indels 41; Gaps 7

OY      8 PPOPTPYPPVPPCPNPVYPQTLHLPOAPPPYDPADPAPVSELYRPSFVIHGPATVPMTSA---- 63
        | :| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      298 PVPEPSPVOPPALPLPASPSVLPAESPSPSLSPSPAFA-----ASMHPLSAAVPAATAPPCA 353
        | :| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      64 -----AFPGLSXYLPM-AQSVAVGLGSTRIPAAV-----YVPVP 96
        | :| ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      354 TVAGASPAVVSSILAMPHDGYLLPKDAFFSLGLASRSRAVPMVPGCAVAAPPSPAPPLPI-P 412
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      97 IYPP--GSTFLVEGGYD 111
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      413 SYPASVGAAPV---GYD 426

RESULT 12
RPBI_DICDI
ID RPBI_DICDI STANDARD: PRt: 902 AA.
AC P35084:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)
GN (FRAGMENT).
NC RPOA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostelida; Dictyosteliom.
RN [1]
RX MEDLINE=93129425; PubMed=1482555;
RA Lam T.Y., Chan L., Yip P., Situ C.H.;
RT "The largest subunit of RNA polymerase II in Dictyostelium: conservation of the unique tail domain and gene expression.";
RL Blochem. Cell Biol. 70:792-799(1992).
CC -I FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.
CC CC
CC -I CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE + RNA(N).
CC CC
CC -I SUBCELLULAR LOCATION: NUCLEAR.
CC CC
CC -I PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC CC
CC -I THE PHOSPHORYLATION ACTIVATES POL2.
CC CC
CC -I MISCELLANEOUS: THREE DISTINCT RNA POLYMERASES WERE FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S AND 28S GENES.
CC CC
CC -I SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S52651; AAB24966.1; -.
CC DICTYDB; DD01054; RPOA.
DR INTERPRO; IPRO00684; -.
DR INTERPRO; IPRO02879; -.
DR PFAM; PF01854; RNA_POL_A2; 1.
DR PROSITE; PS00115; RNA_POL_II_REPEAT; 22.
```

KW Transferase: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT NON\_TER 1  
 FT DOMAIN 714 888 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 SO SEQUENCE 902 AA; 100176 MW; 201D02470322203C CRC64;

Query Match 16.5%; Score 107.5; DB 1; Length 902;  
 Best Local Similarity 29.3%; Pred. No. 0.24;  
 Matches 34; Conservative 12; Mismatches 53; Indels 17; Gaps 5;

OY 1 MNSKGOY-PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYR 47  
 DB 704 MNSKGOY-PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYR 47  
 DB 704 MNSKGOY-PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYR 47

OY 48 PSEVHGAAVPTMSAFAFGASLYLPMASVA-VGPLGSTITMAYTPVGPPIPPGS 102  
 DB 761 PSEVHGAAVPTMSAFAFGASLYLPMASVA-VGPLGSTITMAYTPVGPPIPPGS 102

RESULT 13  
 ID PRFL\_LYCES STANDARD; PRT; 1752 AA.  
 AC P36594;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)  
 DE (RNA POLYMERASE II SUBUNIT 1).  
 GN RPB1.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OS Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;  
 CC Schizosaccharomycetaceae; Schizosaccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RX MEDLINE=91187661; PubMed=2011520;  
 RA Azuma Y., Yamauchi M., Ueshima R., Ishihama A.;  
 RT "Cloning and sequence determination of the Schizosaccharomyces pombe  
 rpb1 gene encoding the largest subunit of RNA polymerase II.";  
 RL Nucleic Acids Res. 19:461-468(1991).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +  
 RNA(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 10 DIFFERENT SUBUNITS.  
 CC -1- THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- PUT: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: X56564; CAA39916.1;  
 CC PIR: S26849; S26849.  
 CC INTERPRO: IPR000684;  
 CC INTERPRO: IPR000722;  
 CC INTERPRO: IPR002879;  
 CC PFM: PF01854; RNA\_POL\_A2; 1.  
 CC PFM: PF00623; RNA\_POL\_A; 1.  
 CC PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 24.

KW Transferase: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 69 85 C2H2-TYPE (POTENTIAL).  
 FT DOMAIN 1554 1752 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 SO SEQUENCE 1752 AA; 194161 MW; 15A4F0B59E60570 CRC64;

Query Match 16.4%; Score 107; DB 1; Length 1752;  
 Best Local Similarity 31.4%; Pred. No. 0.49;  
 Matches 33; Conservative 13; Mismatches 53; Indels 6; Gaps 5;

OY 1 MNSKGOY-PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYRPFVHGAAV 58  
 DB 1572 MNSKGOY-PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYRPFVHGAAV 58  
 DB 1572 MNSKGOY-PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYRPFVHGAAV 58

OY 59 PTMSAFAFGASLYLPMASVA-VGPLGSTITMAYTPVGPPIPPGS 102  
 DB 1629 PTMSAFAFGASLYLPMASVA-VGPLGSTITMAYTPVGPPIPPGS 102

RESULT 14  
 ID PRFL\_LYCES STANDARD; PRT; 346 AA.  
 AC Q00451;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE 36.4 KDA PROLINE-RICH PROTEIN.  
 DE TPRP-F1.  
 GN Lycopersicon esculentum (Tomato).  
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OS Magnoliophyta; Eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Solanales; Solanaceae; Solanum.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. VENT CHERRY; TISSUE=FRUIT;  
 RX MEDLINE=92119262; PubMed=1731999;  
 RA Salts Y., Wachs R., Kenigsbuch D., Grussman W., Barg R.;  
 RT "DNA sequence of the tomato fruit expressed proline-rich protein gene  
 TPRP-F1 reveals an intron within the 3 untranslated transcript.";  
 RL Plant Mol. Biol. 18:407-409(1992).  
 CC -1- FUNCTION: THIS PROTEIN IS A MEMBER OF THE PROLINE-RICH PROTEIN  
 FAMILY. IT IS A MEMBER OF THE PROLINE-RICH PROTEIN FAMILY.  
 CC -1- SUBUNIT: THIS PROTEIN IS A MEMBER OF THE PROLINE-RICH PROTEIN  
 FAMILY. IT IS A MEMBER OF THE PROLINE-RICH PROTEIN FAMILY.  
 CC -1- SUBCELLULAR LOCATION: THIS PROTEIN IS A MEMBER OF THE PROLINE-RICH  
 PROTEIN FAMILY. IT IS A MEMBER OF THE PROLINE-RICH PROTEIN FAMILY.  
 CC -1- PUT: THIS PROTEIN IS A MEMBER OF THE PROLINE-RICH PROTEIN FAMILY.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS A MEMBER OF THE PROLINE-RICH PROTEIN  
 FAMILY. IT IS A MEMBER OF THE PROLINE-RICH PROTEIN FAMILY.  
 CC -1- SIMILARITY: THIS PROTEIN IS A MEMBER OF THE PROLINE-RICH PROTEIN  
 FAMILY. IT IS A MEMBER OF THE PROLINE-RICH PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: X61395; CAA43666.1;  
 CC EMBL: X57076; CAA40361.1;  
 CC HSSP: P24337; IHP.  
 CC INTERPRO: IPR000528;  
 CC PFM: PF00279; LTP; 1.  
 CC PROSITE: PS00279; LTP; 1.  
 SO SEQUENCE 346 AA; 36375 MW; 604E58452FEF16E0 CRC64;

Query Match 15.9%; Score 104; DB 1; Length 346;  
 Best Local Similarity 26.6%; Pred. No. 0.17;  
 Matches 34; Conservative 12; Mismatches 42; Indels 40; Gaps 6;

OY 8 PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYRPFVHGAAV 55  
 DB 66 PTQPTTY-PVOPGPNVYQTLHLPOAPYTDAPAYSELYRPFVHGAAV 55







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:54:50 ; Search time 75.33 Seconds

(without alignments)  
166.711 Million cell updates/sec

Title: US-09-407-430-1-COPY\_1-120

Perfect score: 653  
Sequence: 1 MNSKQYPTQPTYPVQPPGN.....GSTVLVEGVDAARFGAGA 120

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_15:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhcc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_podent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653	100.0	168	4	Q15038
2	648	99.2	168	11	O88675
3	121	18.5	494	10	O99367
4	120.5	18.5	625	10	O99366
5	120.5	18.5	977	10	O99368
6	118	18.1	1840	10	O99368
7	117.5	18.0	1811	5	O99368
8	117.5	18.0	1887	5	O99368
9	117.5	18.0	1889	5	O99368
10	117.5	18.0	1889	5	O99368
11	117.5	18.0	1889	5	O99368
12	116.5	17.8	1553	5	O99368
13	115.5	17.7	1690	5	O99368
14	113.5	17.4	414	3	O99368
15	113	17.3	1605	5	O99368
16	111.5	17.1	237	5	O17242
17	111.5	17.1	1366	11	O08847
18	111.5	17.1	1370	11	O35559
19	111	17.0	1862	5	Q20090

20	110.5	16.9	468	5	Q9NF32	Q9NF32 drosophila
21	110.5	16.9	469	5	Q9W5D6	Q9W5D6 drosophila
22	110.5	16.9	502	4	Q9UG75	Q9UG75 homo sapien
23	107.5	16.5	856	4	Q76022	Q76022 homo sapien
24	105.5	16.2	279	10	Q49490	Q49490 arabidopsis
25	104	15.9	291	10	Q9SK10	Q9SK10 arabidopsis
26	104	15.9	329	4	Q9NR02	Q9NR02 homo sapien
27	103.5	15.8	2457	5	Q77375	Q77375 plasmodium
28	103	15.8	228	5	Q17275	Q17275 brugia paha
29	102.5	15.7	261	4	Q9P112	Q9P112 homo sapien
30	102.5	15.7	530	10	Q9LTK5	Q9LTK5 dictyostell
31	102	15.6	139	5	Q61135	Q61135 dictyostell
32	102	15.6	261	11	P97765	P97765 mus musculu
33	102	15.6	511	5	Q9VXG3	Q9VXG3 drosophila
34	101	15.5	192	5	Q9VXK5	Q9VXK5 drosophila
35	101	15.5	990	4	Q9P257	Q9P257 homo sapien
36	101	15.5	1345	4	Q9P257	Q9P257 caenorhabd
37	100.5	15.4	428	5	Q9P7E8	Q9P7E8 schizosacch
38	100.5	15.4	857	3	Q9VYF9	Q9VYF9 drosophila
39	100	15.3	450	5	Q9VYF9	Q9VYF9 arabidopsis
40	100	15.3	744	10	O65375	O65375 drosophila
41	100	15.3	1039	5	Q9VKG4	Q9VKG4 drosophila
42	99.5	15.2	97	5	Q26291	Q26291 aedes aegypt
43	99.5	15.2	275	10	Q9LFR3	Q9LFR3 arabidopsis
44	99.5	15.2	847	10	Q9X1B6	Q9X1B6 arabidopsis
45	99	15.2	135	5	O61134	O61134 dictyostell

## ALIGNMENTS

RESULT 1  
ID Q15038 PRELIMINARY; PRT; 168 AA.  
AC Q15038;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE KIAA0058 PROTEIN.  
GN KIAA0058.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96051398; PubMed=7584044;  
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,  
RA Seki N., Kawabayashi Y., Ishikawa K., Tabata S.,  
RT "Prediction of the coding sequences of unidentified human genes. II.  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 1:223-229(1994).  
DR EMBL, D31767; BAA06545.1; -;  
DR INTERPRO: IPR000515; -;  
DR PROSITE: PS00402; BPD, TRANSP, INN, MEMBER, UNKNOWN\_1.  
SQ SEQUENCE 168 AA: 17319 MW: 45F1B6D281E24AAC CRG64;

Query Match 100.0%; Score 653; DB 4; Length 168;  
Best local similarity 100.0%; Pred. No. 8.5e-51;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 MNSKQYPTQPTYPVQPPGNPVYQOTLHLPAAPYTDAPPAYSELYRSPFVHPGCAATVPT 60  
DB 1 MNSKQYPTQPTYPVQPPGNPVYQOTLHLPAAPYTDAPPAYSELYRSPFVHPGCAATVPT 60  
QY 61 MSAAPFPGASLTLPMAGSVAAGPLGSTIPMAVYPPGPIPGSTVLVGGVDAARFGAGA 120  
DB 61 MSAAPFPGASLTLPMAGSVAAGPLGSTIPMAVYPPGPIPGSTVLVGGVDAARFGAGA 120  
RESULT 2

088675 PRELIMINARY; PRT; 168 AA.  
AC 088675;  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE PROLINE-RICH PROTEIN.  
GN PRP OR BPR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD1; TISSUE=BRIN;  
RA Yang W., Mansour S.L.;  
RT "A proline-rich protein expressed in mouse brain";  
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF085348; AAC34594.1; -.  
DR MGD; MGI:1344344; Ptbl.  
DR INTERPRO: IPR000515; -.  
DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
SQ SEQUENCE 168 AA; 17288 MW; 49EDDA29D8E344C CRC64;

Query Match 99.2%; Score 648; DB 11; Length 168;  
Best local Similarity 99.2%; Pred. No. 2.4e-50;  
Matches 119; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNSKGYPTPTPTVQPPGPNPVYPTLHLPOAPPTDAPPAVSELYRPSFVHGATVPT 60  
DB 1 MNSKGYPTPTPTVQPPGPNPVYPTLHLPOAPPTDAPPAVSELYRPSFVHGATVPT 60  
QY 61 MSAAPGASLYLPMASVAVGPGSTIPMAVYVPGPIVPGSTVIVEGSDAGARFGACA 120  
DB 61 MSAAPGASLYLPMASVAVGPGSTIPMAVYVPGPIVPGSTVIVEGSDAGARFGACA 120

RESULT 3  
099367 PRELIMINARY; PRT; 494 AA.  
AC 099367;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE II (FRAGMENT).  
GN RPb1-B2.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. WAYNE;  
RA Dietrich M.A.;  
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 204-494 FROM N.A.  
RC STRAIN=CV. WAYNE;  
RX MEDLINE=91355869; PubMed=2103447;  
RA Dietrich M.A., Prenger J., Guilfoyle T.J.;  
RT "Analysis of the genes encoding the largest subunit of RNA polymerase II in Arabidopsis and soybean";  
RL Plant Mol. Biol. 15:207-223(1990).  
DR EMBL; X52493; CAA36734.1; -.  
DR MENDEL; 16056; Glyma; 238; 16056.  
DR INTERPRO: IPR000684; -.  
DR INTERPRO: IPR002879; -.  
DR INTERPRO: IPR002965; -.  
DR PFAM; PF01854; RNA\_POL\_A2; 1.  
DR PRINTS; PRO1217; PRICHEXTENS.  
DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_22.

KW DNA-directed RNA polymerase.  
FT NON\_TER 1  
SQ SEQUENCE 494 AA; 52156 MW; 41EEFD030F61D649 CRC64;

Query Match 18.5%; Score 121; DB 10; Length 494;  
Best local Similarity 32.7%; Pred. No. 0.0033;  
Matches 35; Conservative 11; Mismatches 53; Indels 8; Gaps 4;

QY 3 SKGQYPTPTPTVQPPG----NPVYPT--LHLPOAPPTDAPPAVSELYRPSFVHGAA 56  
DB 242 SPGYSPPTSPYSSPGYSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 300  
QY 57 TPTPMSAAPPASLYLPMASVAVGPGSTIPMAVYVPGPIVPGPS 102  
DB 301 YSPS 347

RESULT 4  
099366 PRELIMINARY; PRT; 625 AA.  
AC 099366;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE (FRAGMENT).  
GN RPb1-B1 OR GENE B1.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. WAYNE;  
RA Dietrich M.A.;  
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 364-625 FROM N.A.  
RC STRAIN=CV. WAYNE;  
RX MEDLINE=91355869; PubMed=2103447;  
RA Dietrich M.A., Prenger J., Guilfoyle T.J.;  
RT "Analysis of the genes encoding the largest subunit of RNA polymerase II in Arabidopsis and soybean";  
RL Plant Mol. Biol. 15:207-223(1990).  
DR EMBL; X52492; CAA36733.1; -.  
DR MENDEL; 16055; Glyma; 238; 16055.  
DR INTERPRO: IPR000684; -.  
DR INTERPRO: IPR002879; -.  
DR INTERPRO: IPR002965; -.  
DR PFAM; PF01854; RNA\_POL\_A2; 1.  
DR PRINTS; PRO1217; PRICHEXTENS.  
DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_23.  
KW DNA-directed RNA polymerase.  
FT NON\_TER 1  
SQ SEQUENCE 625 AA; 68251 MW; DC84628BD8F87AF5 CRC64;

Query Match 18.5%; Score 120.5; DB 10; Length 625;  
Best local Similarity 31.6%; Pred. No. 0.0046;  
Matches 36; Conservative 10; Mismatches 53; Indels 15; Gaps 4;

QY 3 SKGQYPTPTPTVQPPG----NPVYPT-----LHLPOAPPTDAPPAVSELYRPS 49  
DB 395 SPGYSPPTSPYSSPGYSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 453  
QY 50 FVHPGATVPTMSAAPPASLYLPMASVAVGPGSTIPMAVYVPGPIVPGPS 102  
DB 454 YSPS 507

RESULT 5  
099368

```

ID 099368 PRELIMINARY: PRT: 977 AA.
AC 099368:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II (FRAGMENT).
GN RPB1-C.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. WAYNE;
RA Dieckhoff M.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 725-977 FROM N.A.
RC STRAIN-CV. WAYNE;
RA MEDLINE-91355869; PubMed-2103447;
RA Dieckhoff M.A.; Prenger J.; Gullfoyle T.J.;
RT "Analysis of the genes encoding the largest subunit of RNA polymerase
RT II in Arabidopsis and soybean."
RL Plant Mol. Biol. 15:207-223(1990).
DR EMBL: X52495; CAA36736.1;
DR MENDEL: 16057; Glyma:238.16057.
DR INTERPRO: IPR000684;
DR INTERPRO: IPR000722;
DR INTERPRO: IPR002879;
DR PFAM: PF00623; RNA_POL_A: 1;
DR PFAM: PF01854; RNA_POL_A2: 1;
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_17.
KW DNA-directed RNA polymerase.
FT NON_TER 1
SQ SEQUENCE 977 AA; 107938 MW; B25F9AF7EBA0191D CRC64;

Query Match
Best Local Similarity 33.3%; Score 120.5; DB 10; Length 977;
Matches 34; Conservative 12; Mismatches 51; Indels 5; Gaps 4;

OY 3 SKGQYPTQPTVQPPG-NPVYPTQTLHLPOAPYTDAPPAVSELYRSPVHGCAATVPTM 61
DB 756 SPGSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 11
OY 62 SAAFPGASLYLPMAQSVAV-GPLGSTRIPMAVYVPGPIYPPGS 102
DB 813 PSYSPTSPAVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 854

RESULT 6
O99368 PRELIMINARY: PRT: 1840 AA.
AC 099368:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6) II LARGEST CHAIN
DE (EC 2.7.7.6).
GN RPB1.70 OR ATAG35800.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Rose M., Hempel S., Entian K.-D., Hohnselt J., Jesse T.,
RA Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Schueller C.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RA Volckaert G., Grymonprez B., Voet M., Robben J., Mewes H.W.,
RA Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031986; CAA23466.2;
DR EMBL: AL61588; CAB81489.1;
DR INTERPRO: IPR000684;
DR INTERPRO: IPR000722;
DR INTERPRO: IPR002879;
DR INTERPRO: IPR002965;
DR PFAM: PF00623; RNA_POL_A: 1;
DR PFAM: PF01854; RNA_POL_A2: 1;
DR PRINTS: PRO1217; PRICHEXTENS.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_23.
KW DNA-directed RNA polymerase.
SQ SEQUENCE 1840 AA; 204688 MW; 8453621AD945C1B6 CRC64;

Query Match
Best Local Similarity 18.1%; Score 118; DB 10; Length 1840;
Matches 36; Conservative 11; Mismatches 53; Indels 8; Gaps 5;

OY 2 NSKGQYPTQPTVQPPG-NPVYPTQTLHLPOAPYTDAPPAVSELYRSPVHGCA 55
DB 1583 SSPGSGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 11
OY 56 ATVPPTMSAAPGASLYLPMAQSVAV-GPLGSTRIPMAVYVPGPIYPPGS 102
DB 1642 AVSPPTSPAVSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPS 1689

RESULT 7
O99368 PRELIMINARY: PRT: 1811 AA.
AC 099368:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE LARGEST SUBUNIT OF THE RNA POLYMERASE II COMPLEX (FRAGMENT).
GN RPB1215.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7237;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-99242581; PubMed-10224259;
RA Llopert A., Aguade M.;
RT "Synonymous rates at the RPB1215 gene of Drosophila: variation among
RT species and across the coding region."
RL Genetics 152:269-280(1999).
DR EMBL: Y18879; CAB42797.1;
DR INTERPRO: IPR000684;
DR INTERPRO: IPR000722;
DR INTERPRO: IPR002879;
DR INTERPRO: IPR002965;
DR PFAM: PF00623; RNA_POL_A: 1;
DR PFAM: PF01854; RNA_POL_A2: 1;
DR PRINTS: PRO1217; PRICHEXTENS.
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_11.
FT NON_TER 1

```



Query Match 18.0%: Score 117.5; DB 5; Length 1889;  
 Best Local Similarity 30.5%: Pred. No. 0.027;  
 Matches 40; Conservative 10; Mismatches 48; Indels 33; Gaps 7;

QY 8 PTOPTVYVOP---GNPVYPTLHLPOAPPYTDAPPAY---SELYRPSF-----VHP 53  
 DB 1758 PTPSPYSPSPSYDSDGSPGPO---YTPGSPQYSPASPKSPSPPLXSPSSPOHSPSNQYSP 1815  
 QY 54 GAATV-----PTMSAAPGASLYLPMASVA-----VGPLGS-TIPMAVYVGP 96  
 DB 1816 TGSYTSATSPRYSPNMSIYSPSSSTKYSPTPTPTARNYSPTSPMYSPAPSHYSPTSP 1875  
 QY 97 YTPGSTVAVE 107  
 DB 1876 AYSPSSPTFEE 1886

RESULT 10  
 O9XZS1 PRELIMINARY: PRT: 1889 AA.  
 ID O9XZS1  
 AC O9XZS1  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE LARGEST SUBUNIT OF THE RNA POLYMERASE II COMPLEX.  
 GN RPII215.  
 OS Drosophila guanche (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7241;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99242581; PubMed-10224259;  
 RA Llopart A., Aguade M.;  
 RT "Synonymous rates at the RpiI215 gene of Drosophila: variation among  
 RT species and across the coding region.";  
 RL Genetics 152:269-280(1999).  
 DR EMBL: Y18876; CAB42795.1; -;  
 DR INTERPRO: IPR000684; -;  
 DR INTERPRO: IPR000722; -;  
 DR INTERPRO: IPR002879; -;  
 DR INTERPRO: IPR002965; -;  
 DR PFAM: PF00623; RNA\_POL\_A; 1.  
 DR PFAM: PF01854; RNA\_POL\_A2; 1.  
 DR PRINTS: PR01217; PRICHTEXTEN.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_11.  
 SQ SEQUENCE 1889 AA; 209248 MW; D21C6B9B586DD0 CRC64;

Query Match 18.0%: Score 117.5; DB 5; Length 1889;  
 Best Local Similarity 30.5%: Pred. No. 0.027;  
 Matches 40; Conservative 10; Mismatches 48; Indels 33; Gaps 7;

QY 8 PTOPTVYVOP---GNPVYPTLHLPOAPPYTDAPPAY---SELYRPSF-----VHP 53  
 DB 1758 PTPSPYSPSPSYDSDGSPGPO---YTPGSPQYSPASPKSPSPPLXSPSSPOHSPSNQYSP 1815  
 QY 54 GAATV-----PTMSAAPGASLYLPMASVA-----VGPLGS-TIPMAVYVGP 96  
 DB 1816 TGSYTSATSPRYSPNMSIYSPSSSTKYSPTPTPTARNYSPTSPMYSPAPSHYSPTSP 1875  
 QY 97 YTPGSTVAVE 107  
 DB 1876 AYSPSSPTFEE 1886

RESULT 11  
 ID O9XZU9 PRELIMINARY: PRT: 1889 AA.  
 AC O9XZU9;

DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE LARGEST SUBUNIT OF THE RNA POLYMERASE II COMPLEX (RNA POLYMERASE II  
 DE LARGEST SUBUNIT).  
 GN RPII215.  
 OS Drosophila subobscura (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7241;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-RAICES 111;  
 RX MEDLINE-99242581; PubMed-10224259;  
 RA Llopart A., Aguade M.;  
 RT "Synonymous rates at the RpiI215 gene of Drosophila: variation among  
 RT species and across the coding region.";  
 RL Genetics 152:269-280(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-VARIOUS STRAINS;  
 RX MEDLINE-20341326; PubMed-10880485;  
 RA Llopart A., Aguade M.;  
 RT "Nucleotide polymorphism at the RpiI215 gene in drosophila subobscura.  
 RT Weak selection on synonymous mutations.";  
 RL Genetics 155:1245-1252(2000).  
 DR EMBL: Y18876; CAB42800.1; -;  
 DR EMBL: AF272653; AAF89203.1; -;  
 DR EMBL: AF272643; AAF89193.1; -;  
 DR EMBL: AF272644; AAF89194.1; -;  
 DR EMBL: AF272645; AAF89195.1; -;  
 DR EMBL: AF272646; AAF89196.1; -;  
 DR EMBL: AF272647; AAF89197.1; -;  
 DR EMBL: AF272648; AAF89198.1; -;  
 DR EMBL: AF272649; AAF89199.1; -;  
 DR EMBL: AF272650; AAF89200.1; -;  
 DR EMBL: AF272651; AAF89201.1; -;  
 DR EMBL: AF272652; AAF89202.1; -;  
 DR INTERPRO: IPR000684; -;  
 DR INTERPRO: IPR000722; -;  
 DR INTERPRO: IPR002879; -;  
 DR INTERPRO: IPR002965; -;  
 DR PFAM: PF00623; RNA\_POL\_A; 1.  
 DR PFAM: PF01854; RNA\_POL\_A2; 1.  
 DR PRINTS: PR01217; PRICHTEXTEN.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_11.  
 SQ SEQUENCE 1889 AA; 209204 MW; 55D013D56DF82A6B CRC64;

Query Match 18.0%: Score 117.5; DB 5; Length 1889;  
 Best Local Similarity 30.5%: Pred. No. 0.027;  
 Matches 40; Conservative 10; Mismatches 48; Indels 33; Gaps 7;

QY 8 PTOPTVYVOP---GNPVYPTLHLPOAPPYTDAPPAY---SELYRPSF-----VHP 53  
 DB 1758 PTPSPYSPSPSYDSDGSPGPO---YTPGSPQYSPASPKSPSPPLXSPSSPOHSPSNQYSP 1815  
 QY 54 GAATV-----PTMSAAPGASLYLPMASVA-----VGPLGS-TIPMAVYVGP 96  
 DB 1816 TGSYTSATSPRYSPNMSIYSPSSSTKYSPTPTPTARNYSPTSPMYSPAPSHYSPTSP 1875  
 QY 97 YTPGSTVAVE 107  
 DB 1876 AYSPSSPTFEE 1886

RESULT 12  
 ID O96452 PRELIMINARY: PRT: 1553 AA.  
 AC O96452;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DE 01-MAY-2000 (T-EMBLrel. 13, last annotation update)  
DN RNA POLYMERASE II LARGEST SUBUNIT.  
GN RBP1.  
OS Nosema locustae.  
OC Eukaryota; Microsporidia; Nosematidae; Nosema.  
OX NCBI\_TaxID=33696;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99110933; PubMed=9892676;  
RA Hitt R.P., Logsdon J.M. Jr., Healy B., Dorey M.W., Doolittle W.F.,  
RA Embley T.M.,  
RT "Microsporidia are related to Fungi: evidence from the largest subunit  
RT of RNA polymerase II and other proteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:580-585(1999).  
DR EMBL: AF061288; AAD12605.1; -  
DR INTERPRO: IPR000684; -  
DR INTERPRO: IPR000722; -  
DR INTERPRO: IPR002879; -  
DR INTERPRO: IPR002965; -  
DR PFAM: PF00623; RNA\_pol\_A; 1.  
DR PFAM: PF01854; RNA\_pol\_A2; 1.  
DR PRINTS: PR01217; PRICHEXENS.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN 12.  
SQ SEQUENCE 1553 AA; 174466 MW; 309656AB4486F7B8 CRC64;

Query Match	17.8%	Score	116.5	DB	5	Length	1553
Best Local Similarity	31.7%	Pred	No. 0.027				
Matches	32	Mismatches	9	Indels	13	Gaps	4

QY 8 PTOPTV----PVQPPGNPVPYPQT--LHLPLQAAPPYTDAAPAYSELYRPSFVHPCGATVPTM 61  
||| | | : | : | : | : ||| : ||  
Db 1441 PTSPAYSPPTSPAYSPPTSPAYSPPTSPAYSPPTSPAYSPPTSPAYSPPTSPAYSPPTSPAYSPPTSP 1499

```

QY      62  SAAFGASLYLPMAQSVAVGPLGSTITPAAIYYPGIIPPGS 102
          | | | | | | | | | | | | | | | | | | | |
Db    1500  PAYSPPSPAYP-----TSPAYEPTSPAYSPSPVPRPST 1534

```

```

RESULT 13
077165
ID 077165; PRELIMINARY; PRT; 1690 AA.
AC 077165;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE DNA-DEPENDENT RNA POLYMERASE II LARGEST SUBUNIT (FRAGMENT).
GN PRP1.
OS Mastigamoeba invertens.
OC Eukaryota; Mastigamoeba.
OX NCBI_TaxID=81100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50338;
RC MEDLINE=98426228; PubMed=9751740;
RA Stiller J.W., Duffield E.C.S., Hall B.D.;
RT "Amliochondriate amoebae and the evolution of DNA-dependent RNA
RT polymerase II.";
RL Proc. Natl Acad. Sci. U.S.A. 95:11769-11774(1998).
DR EMBL; AF083338; AAC62246.1; -.
DR INTERPRO; IPR007222; -.
DR INTERPRO; IPR002879; -.
DR PFAM; PF00623; RNA_pol_A; 1.
DR PFAM; PF01854; RNA_pol_A2; 1.
FT NON_TER 1
SQ SEQUENCE 1690 AA; 184993 MW; 012CDF9B3DD20F64 CRC64;

```

Query Match	17.7%	Score 115.5;	DB 5;	Length 1690;
Best Local Similarity	31.1%;	Pred. No. 0.037;		
Matches 32;	Conservative	9;	Mismatches 51;	Indels 11;
			Gaps	4
QY	5	GAYPTQPTV----	PVQPRGNPVYRQTLHLRQADPRPYTDAPRAYSSELYRPSFVHPGAAVTVP	60

Db 1505 GYSPASPAIYSPASPAIYSPASPAIS ---- PASPAIYSPASPAISPA -SPAISPAISPAISPA 1555  
 GY 61 MSAAPPGASLYLPMAGSVA -VGPIGSTIPMAYYVPGIYPPGS 102  
 Db 1559 SPAISPAISPAISPAISPAISPAISPAISPAISPAISPAISPAISPAIS 1601

ID	PRELIMINARY;	PRT;	414 AA.
094231			
AC 094231;			
DT 01-MAY-1999 (TrEMBLrel. 10, Created)			
DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)			
DT 01-MAY-2000 (TrEMBLrel. 13, last annotation update)			
DE RNA POLYMERASE II LARGEST SUBUNIT (FRAGMENT).			
OS <i>Kluyveromyces fragilis</i> .			
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales.			
OC Saccharomycetaceae; Kluyveromyces.			
OX NCBI_TaxID=28985;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Lee J.M., Greenleaf A.L.;			
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.			
DR EMBL: U24217; AAC99803.1; -.			
DR INTERPRO: IPR000684; -.			
DR INTERPRO: IPR002879; -.			
DR INTERPRO: IPR002965; -.			
DR PFAM: PF01854; RNA_POL_A2; 1.			
DR PRINTS: PR01217; PRICHEXENS.			
DR PROSITE: PS00115; RNA_POL_II_REPEAT; UNKNOWN_20.			
FT NON_TER 1			
SQ SEQUENCE 414 AA; 43369 MW; 59A11389369B00B3 CRC64;			

Query Match	17.48;	Score 113.5;	DB 3;	Length 414;
Best Local Similarity	32.08;	Pred No 0.012;		
Matches 31;	Conservative	13;	Mismatches 48;	Indels 5;
				Gaps 4;

[illegible]

```

RESULT 15
096446 AC
ID 096446 PRELIMINARY; PRT; 1605 AA.
AC 096446;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE LARGEST SUBUNIT OF RNA POLYMERASE II.
GN RPB1.
OS Valdimorpha necatrix.
OC Eukaryota; Microsporidia; Burenellidae; Valdimorpha.
OX NCBI_TaxID=6039;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99110933; PubMed=9892676;
RA Hirt R.P., Logsdon J.M. Jr., Healy B., Dorey M.W., Doolittle W.F.,
  Embley T.M.;
RT "Microsporidia are related to Fungi: evidence from the largest subunit
  of RNA polymerase II and other proteins."
RL Proc. Natl. Acad. Sci. U.S.A. 96:580-585(1999).
DR EMBL: AF060234; AADI2604.L; -.
DR INTERPRO: IPR000684; -.
DR INTERPRO: IPR000722; -.
DR INTERPRO: IPR002879; -.
DR INTERPRO: IPR002965; -.

```

**Tue Mar 6 12:02:33 2001**

us-09-407-430-1\_copy\_1\_120.rspt

Page 7

DR PFAM: PF00623; RNA\_pol\_A.1.  
DR PFAM: PF01854; RNA\_pol\_A2.1.  
DR PRINTS: PR01217; PRICEXENSN.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_17.  
SO SEQUENCE 1605 AA; 180964 MW; EDCC0410F528BA2C CRC64

Query Match	17.3%	Score 113;	DB 5;	Length 1605;
Best Local Similarity	31.4%;	Pred. No. 0.058;		
Matches 32;	Conservative 12;	Mismatches 50;	Indels 8;	Gaps 4;

QY 8 PTPQPY-----VQDPGPGNPNVPPQ--LHLPDAPYTTAPPAISBELRPSEVHFGAATVPTM 61  
| | | | |  
Db 1476 PTPSPYSPSPSYSPSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSP 1534  
| | | | |  
QY 62 SAAFPGASLYLPMASVYA-VGPIGSLTTPMAATVVCATVPPGS 102  
| | | | |  
Db 1535 PPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSPSYSPSP 1576  
| | | | |

Search completed: March 6, 2001, 11:54:53  
Job time: 250 sec







PR 29-JUL-1994; 94US-0282959.  
XX (CHIR ) CHIRON CORP.  
XX Houghton M, Selby M;  
XX WPI: 1996-129331/13.  
DR N-PSDB: T18246.  
XX  
XX Secretable hepatitis C virus E1 and E2 polypeptide(s) lacking  
PT all/part of the membrane spanning domain - useful in vaccines, and  
PT for diagnostic and therapeutic purposes, e.g. in assays for HCV  
XX  
XX Disclosure; Fig 4a-4c; 46pp; English.  
XX  
XX Truncated versions of the hepatitis C virus E2 envelope protein  
CC (R92935) are obtd. by deletion the membrane-spanning C-terminal  
CC anchor domain (approx. amino acids 337-361) of the protein.  
CC This can be achieved by PCR amplification of the E2 gene (T18246)  
CC using appropriate primers. Truncated versions of the E1  
CC polypeptide (see R92934) are similarly produced. When produced  
CC recombinantly in host cells, the truncated proteins are secreted  
CC into the medium. When co-expressed or combined after separate  
CC expression, the truncated E2 and E1 proteins form a complex. The  
CC truncated proteins and complex are useful in vaccines and for  
CC diagnostic and therapeutic applns.  
XX  
XX Sequence 663 AA:  
SQ  
Query Match 100.0%; Score 2058; DB 17; Length 663;  
Best Local Similarity 100.0%; Pred. No. 4.2e-174;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVGTSGAGHTVSGFVSLAPGAKQNVQLINTGSMHLNSTALNCNDSLTGWLAFY 60  
DB 21 ethvgtsgaghtvsgfvsllapgakqnvqlintgsmhlncnclsintgwlafly 80  
QY 61 HHKFNSSGCEPELASCRLTDFDQGMGPIISYANGSGPDRPCWHPKPGCIYPAKSV 120  
DB 81 hnkfnssgcepelascrltdfdqgmgsipisyangsgpdrpcwlypkpgciypaksv 140  
QY 121 GPVYCFPTSPVVVGTDTDSGAPTYSWGENDVDVFLNTRPPLGNMFCCTMNSTGTFTKV 180  
DB 141 gpyvctfptspvvvgttdtsgaptyswgendvdfvlnttrpplgnwfcctmnsstgftkv 200  
QY 181 CGAPPCVIGAGNNTLHCPPTDCFRKHDPATYSRCGSGPWITPRCLVDYPRYLMHPTIIN 240  
DB 201 cgappcvigagntnlhcptdcfrkhpdatysrcsgspwitrclvdypylmhyptcin 260  
QY 241 YTIETIRMYVGVGVEHRLAEACNMWTEGECDEDRSELSPLLTOTOMOVLPCEFTTLR 300  
DB 261 ytifirmyvgvgehrleaacnmwtgercededrseelsplllttctwqvlpcefttlr 320  
QY 301 ALSTGLIHLHONIYDVOYLXGVSSIASMAIKWEYVVLFLFLADARYCSCLMMMLLSIQ 360  
DB 321 alstglinlhqniydvqyllyvgssiasaalkweyvllflfladarycsclmmmlisq 380  
QY 361 AEA 363  
DB 381 aea 383  
RESULT 2  
W67615  
ID W67615 standard; Protein; 663 AA.  
XX W67615;  
AC  
XX 02-MAR-1999 (first entry)  
DT  
XX Hepatitis C virus E2 protein.  
DE  
XX

KW Hepatitis C virus; E2; HCV; truncation; variant; PCR; amplification;  
KW affinity chromatography; Galanthus nivalis; agglutinin; medicament;  
KW detection; infection.  
XX  
XX Hepatitis C virus.  
OS  
XX  
XX W09850556-A2.  
PN  
XX  
XX 12-NOV-1998.  
PD  
XX  
XX 06-MAY-1998; 98WO-US09097.  
PF  
XX  
XX 06-MAY-1997; 97US-0045675.  
PR  
XX  
XX (CHIR ) CHIRON CORP.  
PA  
XX  
XX Abtignani S, Chien D, Choo QL, Glazer E, Houghton M,  
PI Selby M;  
PI WPI: 1999-034724/03.  
DR N-PSDB: W81371.  
DR  
XX  
XX Methods for isolating truncated HCV E1 and E2 polypeptides - used  
PT in, e.g. immunodiagnostic kits for diagnosis of HCV infection  
PT  
XX  
XX Disclosure; Fig 2A-E; 65pp; English.  
PS  
XX  
XX This sequence represents the Hepatitis C virus (HCV) E2 protein. The  
CC invention relates to the intracellular production and isolation of  
CC C-terminally truncated variants of the E2 protein. The truncations  
CC start from about amino acid residue 500 of the E2 protein and are  
CC generated by PCR amplification of the gene sequence. The truncated  
CC proteins are then produced intracellularly in host cells which are  
CC disrupted. The HCV proteins are then purified preferably by affinity  
CC chromatography, especially using Galanthus nivalis agglutinin resin.  
CC The HCV E1 and HCV E2 polypeptides can be used to manufacture a  
CC medicament useful for detecting the presence or absence of HCV infection  
CC in an individual. They can also be used in a immunodiagnostic test kit  
CC for detecting HCV infection.  
XX  
XX Sequence 663 AA:  
SQ  
Query Match 100.0%; Score 2058; DB 20; Length 663;  
Best Local Similarity 100.0%; Pred. No. 4.2e-174;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVGTSGAGHTVSGFVSLAPGAKQNVQLINTGSMHLNSTALNCNDSLTGWLAFY 60  
DB 21 ethvgtsgaghtvsgfvsllapgakqnvqlintgsmhlncnclsintgwlafly 80  
QY 61 HHKFNSSGCEPELASCRLTDFDQGMGPIISYANGSGPDRPCWHPKPGCIYPAKSV 120  
DB 81 hnkfnssgcepelascrltdfdqgmgsipisyangsgpdrpcwlypkpgciypaksv 140  
QY 121 GPVYCFPTSPVVVGTDTDSGAPTYSWGENDVDVFLNTRPPLGNMFCCTMNSTGTFTKV 180  
DB 141 gpyvctfptspvvvgttdtsgaptyswgendvdfvlnttrpplgnwfcctmnsstgftkv 200  
QY 181 CGAPPCVIGAGNNTLHCPPTDCFRKHDPATYSRCGSGPWITPRCLVDYPRYLMHPTIIN 240  
DB 201 cgappcvigagntnlhcptdcfrkhpdatysrcsgspwitrclvdypylmhyptcin 260  
QY 241 YTIETIRMYVGVGVEHRLAEACNMWTEGECDEDRSELSPLLTOTOMOVLPCEFTTLR 300  
DB 261 ytifirmyvgvgehrleaacnmwtgercededrseelsplllttctwqvlpcefttlr 320  
QY 301 ALSTGLIHLHONIYDVOYLXGVSSIASMAIKWEYVVLFLFLADARYCSCLMMMLLSIQ 360  
DB 321 alstglinlhqniydvqyllyvgssiasaalkweyvllflfladarycsclmmmlisq 380  
QY 361 AEA 363  
DB 381 aea 383

Db 381 aea 383

RESULT 3  
R08123  
ID R08123 standard; protein; 2772 AA.  
XX  
AC R08123;  
XX  
DT 23-JAN-1991 (first entry)  
XX  
DE Hepatitis C virus polypeptide from long ORF.  
XX  
KM Hepatitis C virus; antiviral agent.  
XX  
OS Hepatitis C virus.  
XX  
PN EP388232-A.  
XX  
PD 19-SEP-1990.  
XX  
PF 16-MAR-1990; 90EP-0302866.  
XX  
PR 18-MAY-1989; 89US-0355002.  
PR 17-MAR-1989; 89US-0325338.  
PR 20-APR-1989; 89US-0341334.  
XX  
PA (CHIR-) CHIRON CORP.  
XX  
PI Houghton M, Choo QL, Kuo G;  
XX  
DR WPI; 1990-284418/38.  
DR N-PSDB; 005955.  
XX  
PT Hepatitis C virus DNA - used for producing probes,  
PT polypeptide(s), antibodies and anti-sense polynucleotide(s) for  
PT diagnosis and therapy.  
XX  
PS Disclosure; Fig 16; 83pp; English.  
XX  
CC HCV CDNA libraries were constructed using pooled serum from a  
CC chimpanzee with chronic HCV infection. A lambda gtl1 library was  
CC screened with probes derived from previously isolated clones. The  
CC ORF is derived from the overlapping clones p14a, CA167b, CA156e,  
CC CA84a, CA59a, K9-1, 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36,  
CC 8i, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g and 15e.  
CC This polypeptide can be used to design probes  
CC for the detection of HCV nucleic acids. In screening programmes  
CC for antiviral agents and in preparing blood free of HCV. Antisense  
CC polynucleotides can be used to inhibit viral replication.  
CC See also 005956.  
CC  
XX  
SO Sequence 2772 AA:

Query Match 100.0%; Score 2058; DB 11; Length 2772;  
Best Local Similarity 100.0%; Pred. No. 2,6e-173;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETVVTVGSGAGHTVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSINTGWLGLFY 60  
DB 270 etvvtgsgaghtvsgfvsllapgakonvolintngshlntalnctndstntgwlglfy 329  
QY 61 HHFNSSGCEPERLASCRPLTFDPOGWPISYANGSGPDORPYCMWHPKRCGIVPAKSYC 120  
DB 330 hkfhnssgceperlascrpltdfdogwpslsyansgpdqprycmwhpprcgivpaksvy 389  
QY 121 GPVYCFPPSPVVVGTDRSCAPTYSGENDTDVFLNTRPPLGNFGCTWMSGTGFTKY 180  
DB 390 gpyvcfppspvvvgtdrscaptysgendtdvflntrpplgnfngctwmstgtgftky 449  
QY 181 CGAPPCVIGAGNNTLHCPDRCRKHDPDATYSGSGSPWTPPCLVLDYPRRLMHPCTT 240  
CGAPPCVIGAGNNTLHCPDRCRKHDPDATYSGSGSPWTPPCLVLDYPRRLMHPCTT

Db 450 cgappcvlgagntlnhpcldcfkhpdaelysrgsgpwltprclvdypylwihypcltn 509  
QY 241 YTIKIRMYGCVGVEHRLAECNMWTRGECODEDRSELSPLLTQWVLPDPSFTTTP 300  
DB 510 ytlkirmygcvgvehrleaecnwtrgercdledrseisplllttqvwlpdpsfttclp 569  
QY 301 ALSTGLIHLNONTVDVQYLVGVSSISMAIKMEYVLLFELLADAVSCGLMMMLISO 360  
DB 570 alstgllhlhnlvndvqylvgvsslsaswskwvllflllledavscslmmlisq 629  
QY 361 AEA 363  
DB 630 aea 632

RESULT 4  
ID Y14975  
Y14975 standard; Protein; 2955 AA.  
XX  
AC Y14975;  
XX  
DT 08-NOV-1999 (first entry)  
XX  
DE Amino acid sequence of HCV-1 ORF.  
XX  
KM Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;  
KM HCV infection; vaccine.  
XX  
OS Hepatitis C virus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 441 /note= "encoded by Y1"  
FT Misc-difference 461 /note= "encoded by Y1"  
FT Misc-difference 461 /note= "encoded by CCCC"  
XX  
PN EP939128-A2.  
XX  
PD 01-SEP-1999.  
XX  
PF 17-SEP-1990; 90EP-0310149.  
XX  
PR 21-DEC-1989; 89US-0456142.  
PR 15-SEP-1989; 89US-0408045.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (OYAA/) OYA A.  
PI Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;  
PI Miyamura T, Saito I, Weiner AJ;  
XX  
DR WPI; 1999-480843/41.  
DR N-PSDB; 207656.  
XX  
PT New Hepatitis C Virus isolates, useful for diagnosis of hepatitis  
PT infections and development of vaccines  
XX  
PS Disclosure; Fig 12; 132pp; English.  
XX  
CC The invention provides two new isolates of hepatitis C virus (HCV), J1  
CC and J7. These two isolates comprise nucleotide and amino acid sequences  
CC that are distinct from the HCV isolate HCV-1. The nucleotide sequences  
CC may be used to detect non-A, non-B HCV (NANBH) polynucleotides by  
CC hybridisation for diagnosis of NANBH infections. They may also be used to  
CC screen blood donors, donated blood and blood products for this infection.  
CC The isolates may also be used to isolate other naturally occurring  
CC variants of the virus. The polypeptides may be used as a vaccine for  
CC administration to patients to protect against infection with NANBH. The  
CC present sequence represents the amino acid sequence of HCV-1 ORF.  
XX  
SO Sequence 2955 AA:

```

Query Match      100.0%; Score 2058; DB 20; Length 2955;
Best Local Similarity 100.0%; Pred. No. 2.8e-173;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ETHVYTGSGAGHTVSGFVSLAAGAKONVOLINTNGSMHINSTALNCNDSLNTGMLAGLFY 60
    |||||||
Db 384 ethvgsaghtvsgfvsllapgakvnglntngswlnstalnconsintgwlaaglfy 443
OY 61 HHKFNSSGCPERLASCRPLTDFDGCWGPISYANGSGPDORPCWHPKPGCIVPAKSYC 120
    |||||||
Db 444 hnkfnssgcpelascrpltdfdgwwpisyangsgpdqrpvcwppkpcgitypaksvc 503
OY 121 GPVYCFPTSPVYVGTDDSGAPTYSMGENDTDFVLNTRPRLGNMFGCTWNNSTGFTKV 180
    |||||||
Db 504 gpvyctfptspvvvgltfsgapctysgendtclvlynmtrpplgnwfgctwmnstgftkv 563
OY 181 CGAPPCVIGAGANNLHCPDPCFRKHPDAPYSRCSGPMTPRCLVDYPRYLMHYPCTIN 240
    |||||||
Db 564 cgappcviggagannlhcpdpcfrkhpdatyrcsggpmtprcldvpyrlmhyptcin 623
OY 241 YTFKIRMYGVGEHRLERACNWTGRCDELEDRDSELSPLLTQTQOVLPSCFTTLF 300
    |||||||
Db 624 ytfkirmvgvgehrleacnwtgercdledrdrseplslltqtqvlpstfllp 683
OY 301 AASTGLIHQKIVDVOYLYGVSSIASWAIKWEYVLLFLDLADARVCSCLMMMLISQ 360
    |||||||
Db 684 aistglilhqnivdvqlylygvssiaswakweyvvllflldarvcsclmmmlisq 743
OY 361 AEA 363
    |||
Db 744 aea 746

RESULT 5
R21519
ID R21519 standard; Protein: 3011 AA.
XX AC R21519;
XX DT 22-JUN-1992 (first entry)
XX DE Compiled HCV sequence.
XX KW HCV1; serum; gtl1.
XX OS Hepatitis C virus 1.
XX FH Key Location/Qualifiers
    FH Misc-difference 9 /label= ARG
    FT Misc-difference 11 /label= THR
    FT Misc-difference 176 /label= THR
    FT Misc-difference 334 /label= VAL
    FT Misc-difference 603 /label= ILE
    FT Misc-difference 848 /label= (ASN)
    FT Misc-difference 1114 /label= SER
    FT Misc-difference 1117 /label= THR
    FT Misc-difference 1276 /label= LEU
    FT Misc-difference 1328 /label= (VAL)
    FT Misc-difference 1454 /label= TYR
    FT Misc-difference 1471 /label= (SER)
    FT Misc-difference 1877

```

```

FT /label= (GLY)
FT Misc-difference 1948 /label= (HIS)
FT Misc-difference 1949 /label= (CYS)
FT Misc-difference 2021 /label= (VAL)
FT Misc-difference 2349 /label= (SER)
FT Misc-difference 2385 /label= (PHE)
FT Misc-difference 2386 /label= (ALA)
FT Misc-difference 2502 /label= (PHE)
FT Misc-difference 2690 /label= (GLY)
FT Misc-difference 2996 /label= (PRO)
XX W09202642-A.
XX 20-FEB-1992.
XX 12-AUG-1991; 91WO-US05728.
XX 10-AUG-1990; 90US-0566209.
XX (CHIR-) CHIRON CORP.
XX Houghton M, Choo QL, Kuo G, Weiner AJ, Urdea MS, Irvine BD;
XX Kolberg JA;
XX WPI: 1992-080094/10.
XX N-PSDB: Q21744.
XX Reagents for isolating, amplifying and detecting HCV
XX polynucleotide(s) - used to monitor spread of blood-borne non-a,
XX non-b hepatitis virus infection and screen blood samples for
XX virus
XX Disclosure: Fig 1; 67pp; English.
XX Heterogeneities in cloned DNAs of HCV1 are indicated by the amino
XX acid indicated in the features, the parentheses indicated that the
XX heterogeneity was detected at or near to the 5'- or 3'-end of the
XX HCV in the clone.
XX The sequence is derived from a composite HCV cDNA from HCV1, a
XX prototypic HCV. The DNA sequence is based upon sequence information
XX derived from a no. of HCV cDNA clones, which were isolated from a no.
XX of HCV cDNA libraries, including the "c" library present in lambda
XX gtl1 (ATCC No.40394), and from human serum. The HCV cDNA clones
XX were isolated by methods described in W09014436.
XX The clones from which the sequence was derived are 5' clone32,
XX b114a, 189, aq30a, CA205a, CA290a, CA216a, p114a, CA167b, CA156e,
XX CA84a, CA59a, K9-1 (also called K9-1), 26j, 13i, 12f, 14i, 11b, 7f,
XX 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g,
XX 39c, 35f, 19g, 26g, 15e, b5a, 16jh, 6k and p131jh.
XX Sequence 3011 AA;

Query Match      100.0%; Score 2058; DB 13; Length 3011;
Best Local Similarity 100.0%; Pred. No. 2.9e-173;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ETHVYTGSGAGHTVSGFVSLAAGAKONVOLINTNGSMHINSTALNCNDSLNTGMLAGLFY 60
    |||||||
Db 384 ethvgsaghtvsgfvsllapgakvnglntngswlnstalnconsintgwlaaglfy 443
OY 61 HHKFNSSGCPERLASCRPLTDFDGCWGPISYANGSGPDORPCWHPKPGCIVPAKSYC 120
    |||||||
Db 444 hnkfnssgcpelascrpltdfdgwwpisyangsgpdqrpvcwppkpcgitypaksvc 503

```

QY 121 GPVYCTPSPVVVGTDRSGAPTYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180  
 DB 504 GPVYCTPSPVVVGTDRSGAPTYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 563  
 QY 181 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCTIN 240  
 DB 564 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCTIN 623  
 QY 241 YTIKIRMYVGGVEHRLAEACNWRGRCDEDRSELSPLLLTTTQWQVLPCSFPTLP 300  
 DB 624 YTIKIRMYVGGVEHRLAEACNWRGRCDEDRSELSPLLLTTTQWQVLPCSFPTLP 683  
 QY 301 ALSTGLIHQNIQVVOYLYGVGSSIASWAIKWEYVLLFLLADARVCSCLWMLLSQ 360  
 DB 684 ALSTGLIHQNIQVVOYLYGVGSSIASWAIKWEYVLLFLLADARVCSCLWMLLSQ 743  
 QY 361 AEA 363  
 DB 744 aea 746  
 RESULT 6  
 R09031 ID R09031 standard; Protein: 3011 AA.  
 XX AC R09031;  
 XX DT 15-MAY-1996 (first entry)  
 XX Hepatitis C virus polyprotein.  
 DE Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection;  
 KW diagnosis; antibodies.  
 XX Hepatitis C virus.  
 XX Key Location/Qualifiers  
 FT Misc-difference 1..122  
 FT /label= antigen  
 FT /note= "C22; R09036"  
 FT Misc-difference 199..328  
 FT /label= antigen  
 FT /note= "S2; R09035"  
 FT Misc-difference 1192..1457  
 FT /label= antigen  
 FT /note= "C33c; R09032"  
 FT Misc-difference 1569..1931  
 FT /label= antigen  
 FT /note= "C100; R09033"  
 FT Misc-difference 2054..2464  
 FT /label= antigen  
 FT /note= "NS5; R09034"  
 XX EP693687-A1.  
 XX 24-JAN-1996.  
 XX 03-APR-1991; 91EP-0114016.  
 XX 04-APR-1990; 90US-0504352.  
 XX (CHIR ) CHIRON CORP.  
 XX Choo Q, Houghton M, Kuo G;  
 XX WPI; 1996-117956/13.  
 XX N-PSDB; T12710.  
 XX Combinations of synthetic Hepatitis C Virus antigens - provide more  
 PT effective diagnosis of Non-A, Non-B Hepatitis  
 XX  
 PS Disclosure; Fig 1(A-Y); 53pp; English.

XX The combination comprises an HCV antigen from the C domain (pref.  
 CC C22 - R09036) and at least one HCV antigen from the NS3 (pref. C33c  
 CC - R09032), NS4 (pref. C100 - R09033), S (pref. S2 - R09035) or NS5  
 CC (R09034) domain.  
 CC The antigens may in the form of a fusion protein, a simple physical  
 CC mixture, or the individual antigens commonly bound to a solid matrix.  
 CC They are pref. prepd. by recombinant DNA techniques (primers are  
 CC given in T12711-T12716), but can be synthesised or isolated from  
 CC HCV using affinity chromatography.  
 XX SQ Sequence 3011 AA;  
 Query Match 100.0%; Score 2058; DB 17; Length 3011;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-173;  
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLNSTALNCNDSINTGWLGLFY 60  
 DB 384 ethvtgsgaghtvsgfvsllapgakgnvqlintngshlntalncndsintgwlglfy 443  
 QY 61 HHKFNSSGCPERLASCRPLTDFDGGWCPISYANGSGDPDQRPYCHWHYPKPGIVPAKSVC 120  
 DB 444 hhkfnssgcpelrascrpltdfdggwcpisyangsgdpdqrpychwhypkpgivpaksvc 503  
 QY 121 GPVYCTPSPVVVGTDRSGAPTYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180  
 DB 504 GPVYCTPSPVVVGTDRSGAPTYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 563  
 QY 181 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCTIN 240  
 DB 564 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCTIN 623  
 QY 241 YTIKIRMYVGGVEHRLAEACNWRGRCDEDRSELSPLLLTTTQWQVLPCSFPTLP 300  
 DB 624 YTIKIRMYVGGVEHRLAEACNWRGRCDEDRSELSPLLLTTTQWQVLPCSFPTLP 683  
 QY 301 ALSTGLIHQNIQVVOYLYGVGSSIASWAIKWEYVLLFLLADARVCSCLWMLLSQ 360  
 DB 684 alstglilhqnivdvqylyvgvssiaswaikweyvllflladarcvscclwmlllsq 743  
 QY 361 AEA 363  
 DB 744 aea 746  
 RESULT 7  
 W34480 ID W34480 standard; Protein: 3011 AA.  
 XX AC W34480;  
 XX DT 16-MAR-1998 (first entry)  
 XX HCV polyprotein.  
 DE PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;  
 KW C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;  
 KW NS4.  
 XX Hepatitis C virus.  
 XX Key Location/Qualifiers  
 FH Misc-difference 366  
 FT /note= "can optionally be Arg"  
 FT Misc-difference 372  
 FT /note= "can optionally be Thr"  
 FT Misc-difference 867  
 FT /note= "can optionally be Thr"  
 FT Misc-difference 1341  
 FT /note= "can optionally be Val"  
 FT Misc-difference 2148

FT Misc-difference 2883 /note= "can optionally be Ile"  
 FT Misc-difference 3681 /note= "can optionally be Asn"  
 FT Misc-difference 3681 /note= "can optionally be Ser"  
 FT Misc-difference 3690 /note= "can optionally be Thr"  
 FT Misc-difference 4167 /note= "can optionally be Leu"  
 FT Misc-difference 4323 /note= "can optionally be Val"  
 FT Misc-difference 4701 /note= "can optionally be Tyr"  
 FT Misc-difference 4752 /note= "can optionally be Ser"  
 FT Misc-difference 5970 /note= "can optionally be Gly"  
 FT Misc-difference 6183 /note= "can optionally be His"  
 FT Misc-difference 6186 /note= "can optionally be Cys"  
 FT Misc-difference 6402 /note= "can optionally be Val"  
 FT Misc-difference 7386 /note= "can optionally be Ser"  
 FT Misc-difference 7494 /note= "can optionally be Phe"  
 FT Misc-difference 7497 /note= "can optionally be Ala"  
 FT Misc-difference 7845 /note= "can optionally be Phe"  
 FT Misc-difference 8409 /note= "can optionally be Gly"  
 FT Misc-difference 9102 /note= "can optionally be Gly"  
 FT Misc-difference 9327 /note= "can optionally be Pro"

US5683864-A.

04-NOV-1997.

07-JUL-1992; 92US-0910760.

07-JUL-1992; 92US-0910760.

18-NOV-1987; 87US-0122714.

30-DEC-1987; 87US-0139886.

26-FEB-1988; 88US-0161072.

06-MAY-1988; 88US-0191263.

26-OCT-1988; 88US-0263584.

14-NOV-1988; 88US-0271450.

17-MAR-1989; 89US-0325338.

20-APR-1989; 89US-0341334.

21-APR-1989; 89US-0353896.

04-APR-1990; 90US-0504352.

(CHTR ) CHIRON CORP.

Choo Q, Houghton M, Kuo G;

WPI; 1997-548976/50.

N-PSDB; T99981.

Combination of three hepatitis C virus antigens - used for detection of specific antibodies to diagnose infection

Disclosure; Column 25-46; 57pp; English.

XX This sequence represents the Hepatitis C virus polyprotein. Fragments of  
 CC the DNA encoding this sequence can be amplified and used in the  
 CC combination of HCV antigens of the invention. The HCV antigen combination  
 CC comprises an antigen (Agl) comprising the C domain (i.e. amino acids (aa)  
 CC 1-120 of the HCV polyprotein), or its immunologically reactive fragment

CC containing at least 8 aa. It also comprises two additional antigens from  
 CC two different polyprotein domains, including at least 8 aa from the NS3,  
 CC NS4, S or NS5 domains of the polyprotein, corresponding, respectively, to  
 CC aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein.  
 CC Alternatively, Agl contains at least 8 aa from the 1-122 or 9-177 aa  
 CC regions of the HCV polyprotein. These antigen combinations are used  
 CC diagnostically to detect anti-HCV antibodies, using any standard  
 CC immunoassay format. These antigen combinations have a broader range of  
 CC reactivity with antibodies than any antigen individually.

XX Sequence 3011 AA;

Query Match 100.0%; Score 2058; DB 18; Length 3011;

Best Local Similarity 100.0%; Pred. No. 2.9e-173;

Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 ETHVTGSGAGTYSGVFVLLAPGAKQNVOLINTNGSHLNSALNCNDSINTGWLGLFY 60  
 Db 384 ethvtgsgaghtvsgfvslilapgaknqvlnltngshlnstalnclndsintgwlglfy 443  
 QY 61 HKFNSGCGPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCWHPKPCGIVPAKSV 120  
 Db 444 hkfnsngcperlasrpltdfdogwgpisyangsgdpqrpypcwppkpcgivpaksv 503  
 QY 121 GPVYCFPTSPVVGTTDRSGAPTYSWGENDTDFVLNTRPPPLGNWFGCTWMSSTGFTKV 180  
 Db 504 gpvycftpspvvgtttdrsgaptyswgendtdfvlntntrppplgnwfgctwmsstgftkv 563  
 QY 181 CGAPPCVTGGAGNNTLHCPTDCFRKHDPATYSRGSGGPWTTPRCLVDYPPYRLWHYPC 240  
 Db 564 cgappcvtgagntlhcptdcfrkhdpdatsrgsggptwtprrclvdyppyrwhypc 623  
 QY 241 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWVLPSCFTTLP 300  
 Db 624 ytifkirmvgygvehrleaecnwtgercdledrdrselpllltttqwvlpcsfttlp 683  
 QY 301 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVLLFLLLDARVCSCLWMLLSQ 360  
 Db 684 alstglihlhqnivdvqylyvgvssiaswaikweyvllfllldarvcscilwmlllsq 743  
 QY 361 AEA 363  
 Db 744 aea 746

RESULT 8

W40038

ID W40038 standard; Protein; 3011 AA.

XX AC W40038;

XX 26-MAY-1998 (first entry)

XX HCV polyprotein.

XX Hepatitis C virus C domain; HCV; C antigen; immunological activity;

KW NS3 domain; NS4 domain; S domain; NS5 domain.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT Domain 1..120

FT Modified-site 9 /label= C\_domain

FT /note= "As given in the specification this amino acid can also be Arg"

FT Modified-site 11

FT /note= "As given in the specification this amino acid can also be Thr"

FT Domain 120..400

FT Modified-site 174 /label= S\_domain

FT FT /note= "As given in the specification this amino  
 Modified-site 334 acid can also be Thr"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 603 acid can also be Val"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 847 acid can also be Ile"  
 FT FT /note= "As given in the specification this amino  
 Domain 1050..1640 acid can also be Asn"  
 FT FT /label= NS3\_domain  
 Modified-site 1114  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1217 acid can also be Ser"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1276 acid can also be Thr"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1328 acid can also be Leu"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1452 acid can also be Val"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1472 acid can also be Tyr"  
 FT FT /note= "As given in the specification this amino  
 Domain 1640..2000 acid can also be Ser"  
 FT FT /label= NS4\_domain  
 Modified-site 1877  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1948 acid can also be Gly"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 1949 acid can also be His"  
 FT FT /note= "As given in the specification this amino  
 Domain 2000..3011 acid can also be Cys"  
 FT FT /label= NS5\_domain  
 Modified-site 2021  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2348 acid can also be Val"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2385 acid can also be Ser"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2386 acid can also be Phe"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2502 acid can also be Ala"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2690 acid can also be Phe"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2921 acid can also be Gly"  
 FT FT /note= "As given in the specification this amino  
 Modified-site 2996 acid can also be Gly"  
 FT FT /note= "As given in the specification this amino  
 acid can also be Pro"

US5712087-A.

27-JAN-1998.

PF 12-MAY-1995; 95US-0440519.  
 XX  
 PR 07-JUL-1992; 92US-0910760.  
 PR 04-APR-1990; 90US-0504352.  
 PR 12-MAY-1995; 95US-0440519.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Choo Q, Houghton M, Kuo G;  
 XX  
 DR WPI; 1998-119973/11.  
 DR N-PSDB; V09989.  
 XX  
 PT Immunoassays for hepatitis C virus antibodies - using combinations  
 PT of antigenic fragments of HCV polyprotein  
 XX  
 PS Disclosure; Fig 1; 59pp; English.  
 XX  
 CC This sequence represents the hepatitis C virus (HCV) polyprotein which  
 CC is used in the construction of novel combinations of HCV antigens that  
 CC have a broader range of immunological activity than any single HCV  
 CC antigen. An example of such an antigen given in this specification  
 CC comprises a first antigen containing at least 8 amino acids of the  
 CC C domain of the HCV polyprotein and a second antigen comprising at least  
 CC 8 amino acids of the NS3 domain, the NS4 domain, the S domain or the NS  
 CC domain of the HCV polyprotein in the form of a fusion protein, a physical  
 CC mixture or bound to a solid matrix.  
 CC Note: The features given in the specification as represented in the  
 CC feature table of W40038 differ from the positions indicated in Figure 1.  
 XX  
 SQ Sequence 3011 AA;

Query Match 100.0%; Score 2058; DB 19; Length 3011;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-173;  
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHVTVGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 30  
 DB 384 EHVTVGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 30  
 QY 61 HUKNSSGGPERLASCRPLTDFDQGWGPISYANGSGDORPCYWHYPPKCGIVPAKSV 120  
 DB 444 HUKNSSGGPERLASCRPLTDFDQGWGPISYANGSGDORPCYWHYPPKCGIVPAKSV 120  
 QY 121 GPVYCTPSPVVVGTTRDSGAPTSYWGENDTDFVLNNTTRPPLGNWFGTWMNSTGFTKV 180  
 DB 504 GPVYCTPSPVVVGTTRDSGAPTSYWGENDTDFVLNNTTRPPLGNWFGTWMNSTGFTKV 180  
 QY 181 CGAPCVGGAGNNTLHCPDTCFRKHPDATYSRGSGPMTPRCLVDYPRYLWHYPCTIN 240  
 DB 564 CGAPCVGGAGNNTLHCPDTCFRKHPDATYSRGSGPMTPRCLVDYPRYLWHYPCTIN 240  
 QY 241 YTFIKRMVVGVEHRLAECNWTGRCGLDRLDRSELSPILLTTTOMQVLPCSFSTTLP 300  
 DB 624 YTFIKRMVVGVEHRLAECNWTGRCGLDRLDRSELSPILLTTTOMQVLPCSFSTTLP 300  
 QY 301 ALSTGLIHQHNIYDVQVLYGVGSSIASWAIKWEYVLLFLLLADARVCSCLMMMLLSQ 360  
 DB 684 ALSTGLIHQHNIYDVQVLYGVGSSIASWAIKWEYVLLFLLLADARVCSCLMMMLLSQ 360  
 QY 361 AEA 363  
 DB 744 AEA 746

RESULT 9  
 R33992  
 ID R33992 standard; Protein: 480 AA.  
 XX  
 AC R33992;  
 XX  
 DT 26-JUL-1993 (first entry)

XX DE HCV-1 E2/NS1 protein.  
XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;  
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;  
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;  
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.  
XX OS Synthetic.  
XX PN WO9306126-A.  
XX PD 01-APR-1993.  
XX PF 11-SEP-1992; 92WO-US07683.  
XX PR 13-SEP-1991; 91US-0759575.  
XX PA (CHIR ) CHIRON CORP.  
XX PI Houghton M, Weiner AJ;  
XX WPI; 1993-117468/14.  
XX DR  
XX OS Immuno-reactive hepatitis C virus polypeptide compsns. - contg.  
XX PT at least 2 sequences from the first variable domain of distinct  
XX PT HCV isolates  
XX PS Disclosure; Fig 3; 106pp; English.  
XX  
XX The sequences given in R33992-002 represent a portion of the E2/NS1  
XX protein encoded by group I and group II HCV isolates, from amino acid  
XX 370-850. E2/NS1 are viral envelope proteins and are of immunogenic  
XX interest. E2/NS1 contains an N-terminal hypervariable domain of about  
XX 30 amino acids which shows large variation between nearly all isolates.  
XX This is an important immunoreactive domain. This putative envelope  
XX glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera  
XX virus) envelope polypeptide of the pestiviruses and the NS1 of the  
XX flaviviruses, both of which confer protective immunity in hosts  
XX vaccinated with these polypeptides. It has been discovered that a  
XX number of important HCV epitopes vary among viral isolates and that  
XX these epitopes can be mapped to specific domains. This meant that  
XX immunologically cross-reactive polypeptides which focus on variable  
XX rather than constant domains can be produced. See also Q39134-48  
XX and R33982-91.  
XX SQ Sequence 480 AA;  
Query Match 99.9%; Score 2056; DB 14; Length 480;  
Best Local Similarity 99.7%; Pred. No. 4.1e-174;  
Matches 362; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGLAGLFY 60  
DB 15 ethvtgsgaghtvsgfvsllapgakvnlntngswhlntstalnclndslntglaglfy 74  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPIYSANGSGPDORPYCWHYPKPCGIVPAKSV 120  
DB 75 hkhfnssgcpelarscrpltdfdqgwpisyangsgpdqrpvcwhypkpcgivpaksvc 134  
QY 121 GPVYCFTSPVVVGTDRSGAPTSYSGENDTDVFLNTRPPLGNWFGCTWNNSTGFTKV 180  
DB 135 gpvycftspvvvgttdrsgaptsyngendtdvflntrpplgnwfgctwmnstgftkv 194  
QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGGPWITPRCLVDYPYRLWHYPCTIN 240  
DB 195 cgappcvigagntlhcptdcfrkhpdatyrcsgsggpwitrclvdypyrllwhypctin 254  
QY 241 YTIKIRMYVGVGVEHLEAACNWTGERCDLEDNRSELSPLLLTTTQWVLPSCFTTLP 300  
DB 255 ytikirmyvgvgehrleaacnwtgercdlednrseelspllltttqvwlpscfttlp 314

QY 301 ALSTGLIHLHONIVDVQYLYGVGSSIASWAIKWEYVVLLFELLADARVCSCLMWMLISQ 360  
DB 315 alsstglihlnqivdvqylyvgvssiaswaikweyvvllfilladarcvscclmmllisq 374  
QY 361 AEA 363  
DB 375 aea 377  
RESULT 10  
R34009  
ID R34009 standard; Protein; 2816 AA.  
XX AC R34009;  
XX XX  
XX 26-JUL-1993 (first entry)  
XX DE HCV-1 polypeptide.  
XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;  
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;  
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;  
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.  
XX OS Hepatitis C virus.  
XX PN WO9306126-A.  
XX PD 01-APR-1993.  
XX PF 11-SEP-1992; 92WO-US07683.  
XX PR 13-SEP-1991; 91US-0759575.  
XX PA (CHIR ) CHIRON CORP.  
XX PI Houghton M, Weiner AJ;  
XX WPI; 1993-117468/14.  
XX DR  
XX PT Immuno-reactive hepatitis C virus polypeptide compsns. - contg.  
XX PT at least 2 sequences from the first variable domain of distinct  
XX PT HCV isolates  
XX PS Disclosure; Fig 9; 106pp; English.  
XX  
XX This sequence represents the entire hepatitis C virus polypeptide.  
XX HCV is a member of the flavivirus family and appears to encode a basic  
XX polypeptide domain ("C") at the N-terminal of the viral polypeptide,  
XX followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the  
XX nonstructural genes NS2 through NS5. See also Q39134-48, R33982-  
XX 4008 and R38088-89.  
XX SQ Sequence 2816 AA;  
Query Match 99.68%; Score 2050; DB 14; Length 2816  
Best Local Similarity 99.7%; Pred. No. 1.4e-172;  
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ETHVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGLAGLFY 60  
DB 384 ethvtgsgaghtvsgfvsllapgakvnlntngswhlntstalnclndslntglaglfy 443  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPIYSANGSGPDORPYCWHYPKPCGIVPAKSV 120  
DB 444 hkhfnssgcpelarscrpltdfdqgwpisyangsgpdqrpvcwhypkpcgivpaksvc 503  
QY 121 GPVYCFTSPVVVGTDRSGAPTSYSGENDTDVFLNTRPPLGNWFGCTWNNSTGFTKV 180  
DB 504 gpvycftspvvvgttdrsgaptsyngendtdvflntrpplgnwfgctwmnstgftkv 563  
QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGGPWITPRCLVDYPYRLWHYPCTIN 240



```

|||||
Db 564 cgappcvlggagnttlhptdcfrkhpdatysrcsgpwiprcldvpyrlwhypctin 623
Qy 241 YTIKIRMYVGVGVRLEAACNWTGRGCDLEDRSLSPLLTQWQVLPSCFTTLP 300
Db 624 ytficirmyvgvghrleaacnwtgrgcdledrslspllltttqgwvlpcsfittlp 683
Qy 301 ALSTGLIHLHQINVDVQYLYGVGSSIASWATKWEYVVLFLLLADARVCSCLWMLLSQ 360
Db 684 alstglhlhqnivdvqylvgvssiaswalkweyvvllfilladarcvscilwmmllisq 743
Qy 361 AEA 363
Db 744 aea 746

RESULT 11
R08124
ID R08124 standard; protein; 2955 AA.
XX
AC R08124;
XX
DT 23-JAN-1991 (first entry)
XX
DE Hepatitis C virus putative polyprotein.
XX
KW Hepatitis C virus (HCV); antiviral agent.
XX
OS Hepatitis C virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 9..9 /label=K or R
FT Misc-difference 11..11 /label=N or T
FT Misc-difference 176..176 /label=I or T
FT Misc-difference 334..334 /label=M or V
FT Misc-difference 603..603 /label=I or L
FT Misc-difference 848..848 /label=Y or N
FT Misc-difference 1114..1114 /label=P or S
FT Misc-difference 1117..1117 /label-S or T
FT Misc-difference 1276..1276 /label-P or L
FT Misc-difference 1454..1454 /label-C or Y
FT Misc-difference 1471..1471 /label-T or S
FT Misc-difference 1877..1877 /label-E or G
FT Misc-difference 1948..1948 /label-L or H
FT Misc-difference 1949..1949 /label-S or C
FT Misc-difference 2021..2021 /label-V or G
FT Misc-difference 2349..2349 /label-T or S
FT Misc-difference 2385..2385 /label-Y or F
FT Misc-difference 2386..2386 /label-S or A
FT Misc-difference 2502..2502 /label-L or F
FT Misc-difference 2690..2690 /label-R or G
FT Misc-difference 2921..2921 /label-R or G

```

```

XX EP388232-A.
PN
XX
XX 19-SEP-1990.
XX
XX 16-MAR-1990; 90EP-0302866.
XX
XX 18-MAY-1989; 89US-0355002.
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
XX
XX (CHIR-) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
PI
XX
XX WPI; 1990-284418/38.
DR N-PSDB; Q05956.
XX
XX Hepatitis C virus DNA - used for producing probes,
PT polypeptide(s), antibodies and anti-sense polynucleotide(s) for
PT diagnosis and therapy.
XX
XX Disclosure; Fig 17; 83pp; English.
XX
XX HCV cDNA libraries were constructed using pooled serum from a
CC chimpanzee with chronic HCV infection. A lambda gt11 library was
CC screened with probes derived from previously isolated clones. The
CC ORF is derived from the overlapping clones b114a, ag30a, CA205a,
CC CA290a, CA216a, p14a, CA167b, CA156e, CA84a, CA59a, K9-1, 26f, 131,
CC 12f, 14f, 11b, 7f, 8f, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,
CC 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a and 16jh.
CC Polypeptide encoded by this sequence can be used to design probes
CC for the detection of HCV nucleic acids, in screening programmes
CC for antiviral agents and in preparing blood free of HCV. The
CC sequence contains 188 (overlapping) peptides which are claimed as
CC HCV epitopes.
CC See also Q05955.
XX
XX Sequence 2955 AA;
SQ

```

```

Query Match 99.6%; Score 2050; DB 11; Length 2955;
Best Local Similarity 99.4%; Pred. No. 1.4e-172;
Matches 361; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ETHVTGGSAGHTVSGFVSLLPAGKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFY 64
Db 384 ethvtggsaghtvsgfvsllappgkqvqlintngswhlncnclntgwlglify 443
Qy 61 HHKFNSSGCCPERLASCRPLTDFDQGWGPISYANGSGPOORPYCWHYPPKPGIVPAKSVC 120
Db 444 hkhfnssgccperlasrpltdfdqgwgpisyangsgpdqrpvcwhypkpgivpakavc 503
Qy 121 GPVYCFPTSPVVVGTTRSGAPTYSWGENDTDVFLNTRPPLGNWFCCTWMNSTGFTKV 180
Db 504 gpvyctftspvvvgttrsgaptyswgendtdvflntrpplgnwfcctwmnstgftkv 563
Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSPWITPRCLVDYPPYRLWHYPCPIN 240
Db 564 cgappcviggagnttlhcptdcfrkhpdatysrcsgspwxtprclvdypyrwhypctin 623
Qy 241 YTIKIRMYVGVGVRLEAACNWTGRGCDLEDRSLSPLLTQWQVLPSCFTTLP 300
Db 624 ytfikirmyvgvghrleaacnwtgrgcdledrslspllltttqgwvlpcsfittlp 683
Qy 301 ALSTGLIHLHQINVDVQYLYGVGSSIASWATKWEYVVLFLLLADARVCSCLWMLLSQ 360
Db 684 alstglhlhqnivdvqylvgvssiaswalkweyvvllfilladarcvscilwmmllisq 743
Qy 361 AEA 363
Db 744 aea 746

```



CC The construction of pHCV-92 (full-length NS1) involved using the  
 CC xhoI/BamHI insert from pHCV-78 (see R33590) and ligating that into  
 CC the XhoI/BamHI vector backbone of pHCV-80 (see R33591). The resultant  
 CC HCV gene represents amino acids 365-847 of the HCV genome. This  
 CC resulted in a 1449bp EcoRI/BamHI fragment of HCV cloned into the  
 CC CKS fusion vector pJO200. The complete amino acid sequence of the  
 CC antigen is designated pHCV-92 (i.e. R33592). The resultant fusion  
 CC protein HCV CKS-full length NS1 consists of 239 amino acids of CKS,  
 CC seven amino acids contributed by linker DNA sequences and 483 amino  
 CC acids from the NS1 region of the HCV genome. The fusion protein is  
 CC used to detect antibodies and antigen in body fluids from  
 CC individuals exposed to HCV.

XX SQ Sequence 738 AA;

Query Match 99.0%; Score 2037; DB 14; Length 738;  
 Best Local Similarity 98.1%; Pred. No. 3.5e-172;  
 Matches 362; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ETHVTGSGAGTYSVGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60  
 DB 268 ethvtgsgaghtvsgfvslapgaknvolintngswhlncndsintgwlglfy 327  
 QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCMWHPKPGGIVPAKSV 120  
 DB 328 hkhfnssgcperialascrpltdfdqgwqisyangsgdpqrpypcwhypkpggivpaks 387  
 QY 121 GPVYCTFTSPVVVGTDRSCAPTSYSGENDTDVFLNTRPPLGNWFGCTWMNSTGFTKV 180  
 DB 388 gpyvctftspvvvgttdrsgaptsysgendtdvflntrpplgngwfgctwmnstgftkv 447  
 QY 181 CGA-----PPCVIGGAGNNTLCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWH 234  
 DB 448 cgappcvigagnttlhpcptdcfrkhpdatysrcsgsgpwtprclvdypylwh 507  
 QY 235 YPTINTYTFKIMYVGGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWQVLP 294  
 DB 508 yptintytfkirmyvggvehrleaecnwtgercdledrldrseelspllltttqwqlpc 567  
 QY 295 SFTTLPALSTGLHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWM 354  
 DB 568 sfttlpalstglhlhqnivdvqylyvgvssiaswaikweyvvlfllladarvcsclwm 627  
 QY 355 MLLISQAEA 363  
 DB 628 mllisqaea 636

RESULT 14

R34438  
 ID R34438 standard; Protein; 402 AA.

XX AC R34438;

DT 09-AUG-1993 (first entry)

DE Sequence of glycoprotein E2/NS1 in clone HCV1.

XX Hepatitis C virus; envelope protein; glycoprotein; E2/NS1;  
 KW diagnostic reagent.

OS Hepatitis C virus.

PN EP537626-A.

XX PD 21-APR-1993.

XX PF 08-OCT-1992; 92EP-0117191.

XX PR 08-OCT-1991; 91JP-0260824.

XX PA (NAHE-) NAT INST OF HEALTH.

XX PI Harada S, Honda Y, Miyamura T, Saito I;  
 XX DR WPI; 1993-127516/16.  
 DR N-PSDB; Q40330.

XX Diagnostic reagent for hepatitis C virus - comprises second  
 PT envelope protein or first non-structural protein encoded by HCV  
 PT gene and has sugar chain

XX PS Claim 2; Pages 30-32; 58pp; English.

XX Glycoprotein E2/NS1 is derived from the second envelope protein or  
 CC first non-structural protein encoded by the genome of HCV. The  
 CC nucleic acid is extracted from the serum of the patient of hepatitis  
 CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier  
 CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,  
 CC it is preferred to use polymerase chain reaction method. In the  
 CC reaction, any commercially available random primers or synthesized  
 CC DNA having a base sequence similar to that of primer AS1 may be used  
 CC as a primer. Representative examples of sense primers includes S1.

XX SQ Sequence 402 AA;

Query Match 98.9%; Score 2036; DB 14; Length 402;  
 Best Local Similarity 100.0%; Pred. No. 2e-172;  
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60  
 DB 45 ethvtgsgaghtvsgfvslapgaknvolintngswhlncndsintgwlglfy 104  
 QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCMWHPKPGGIVPAKSV 120  
 DB 105 hkhfnssgcperialascrpltdfdqgwqisyangsgdpqrpypcwhypkpggivpaks 164  
 QY 121 GPVYCTFTSPVVVGTDRSCAPTSYSGENDTDVFLNTRPPLGNWFGCTWMNSTGFTKV 180  
 DB 165 gpyvctftspvvvgttdrsgaptsysgendtdvflntrpplgngwfgctwmnstgftkv 224  
 QY 181 CGAPPVCVIGAGNNTLCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWHYPCTTN 240  
 DB 225 cgappcvigagnttlhpcptdcfrkhpdatysrcsgsgpwtprclvdypylwhypcttin 284  
 QY 241 YTFIKIRMTYVGGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWQVLP 300  
 DB 285 ytfikirmvgyggvehrleaecnwtgercdledrldrseelspllltttqwqlpcsf 344  
 QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMMLLI 358  
 DB 345 alstglhlhqnivdvqylyvgvssiaswaikweyvvlfllladarvcsclwmmlli 402

RESULT 15

R24440  
 ID R24440 standard; protein; 2894 AA.

XX AC R24440;

XX DT 02-DEC-1992 (first entry)

XX Composite HCV HC-J1/CDC/CHI protein.

XX Hepatitis C virus; peptides; antibodies; ELISA.

XX Synthetic.

XX Key Location/Qualifiers

XX Peptide 1..20

XX Peptide /label= 1

XX Peptide 7..26

XX Peptide /label= 2

FT Peptide 13..32  
FT /label= 3  
FT Peptide 37..56  
FT /label= 4  
FT Peptide 49..68  
FT /label= 5  
FT Peptide 61..80  
FT /label= 6  
FT Peptide 73..92  
FT /label= 7  
FT Peptide 168..1707  
FT /label= 8  
FT Peptide 1694..1713  
FT /label= 9  
FT Peptide 1706..1725  
FT /label= 10  
FT Peptide 1712..1731  
FT /label= 11  
FT Peptide 1718..1737  
FT /label= 12  
FT Peptide 1724..1743  
FT /label= 13  
FT Peptide 1730..1749  
FT /label= 14  
FT Peptide 2263..2282  
FT /label= 15  
FT Peptide 2275..2294  
FT /label= 16  
FT Peptide 2287..2306  
FT /label= 17  
FT Peptide 2299..2318  
FT /label= 18  
FT Peptide 2311..2330  
FT /label= 19  
XX  
XX EP489968-A.  
XX  
XX 17-JUN-1992.  
XX  
XX 14-DEC-1990; 90EP-0124241.  
XX  
XX 14-DEC-1990; 90EP-0124241.  
XX (INNO-) INNOGENETICS NV.  
XX  
XX Deleys RJ, Maertens G, Pollet D, Van Heuverswyn H;  
PI WPI; 1992-201383/25.  
XX  
XX New synthetic peptides for detecting antibodies to hepatitis C  
PT virus - useful in e.g. ELISA assays, and for detection of HCV  
PT antigens or as immunogens  
XX  
XX Disclosure; Fig 1; 32pp; English.  
XX  
XX RNA viruses frequently exhibit a high rate of spontaneous mutation,  
CC thus a virus is considered to be the same of equiv. to HCV if it  
CC exhibits a global homology of more than 70 percent with the HCV HC-  
CC J1/CDC/CHI composite sequence. The peptide fragments of this DNA  
CC sequence indicated in the features table can immunologically mimic  
CC proteins encoded by HCV. Additional amino acids or chemical gps.  
CC may be added to either end of the peptides for the purpose of  
CC creating a linker arm for attachment to a carrier. The peptides can  
CC be used for the detection of antibodies specific for HCV. They may  
CC be used in the form of kits, opt. with reagents such as  
CC staphylococcal protein A, streptococcal protein G, avidin or  
CC streptavidin. The peptides may also be used as immunogens for  
CC raising antibodies.  
XX  
XX Sequence 2894 AA;

Query Match 96.3%; Score 1982; DB 13; Length 2894;

Best Local Similarity 95.9%; Pred. No. 1.5e-166;  
Matches 348; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
QY 1 ETHVTGSGAGHTVSGFVSLIAPGAKONVOLINTNGSHLNSTALNCNDSINTGWLGLFY 60  
Db 384 etivsggqaaramsgivslftpgaknqlintngshlnstalnncnesintgwlglly 443  
QY 61 HRFNSGSCPERIASCRPLTDFDQGWGCPISYANGSGPDQRPYCWHPYPPKPCGIVPAASVC 120  
Db 444 qhkfnsngscperlasrpltdfdqggwgpisyangsgpdqrpypcwhppkpcgivpaksvc 503  
QY 121 GPVYCFTSPVVGTTDRSGAPTYSMGENDTDFVVLNTRPPLGNMFGCTWMNSTGFTKV 180  
Db 504 gpvycftspvvvgtttdrsgaptyswgendtdvfvlntrpplgnwfgctwmnstgftkv 563  
QY 181 CGAPPCVIGGAGNNTLHCPDTCPRKHPDATYSRCGSGPWITPRCLVDYDYLWHYPCTIN 240  
Db 564 cgappcviggnntlhcptdcprkhpdatysrcsgspwitrclvdypylwhypctin 623  
QY 241 YTIKIRMYGGVEHRLAEAAACNMTRGERCDLEDRDRSELSPLLLITTTQWVLPSCFTTLP 300  
Db 624 ytifkirmvvggvehrleaaacnmtrgercdledrdrseelspllltttqwgvlpcsfllp 683  
QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSSTASWAKWEYVVLLELLADARVCSCLWMLLIQS 360  
Db 684 alstgllhlhqnivdvqylyvgvssstaswawkweyvvllelladarvcscwmlllisq 743  
QY 361 AEA 363  
Db 744 aea 746

Search completed: March 6, 2001, 11:51:45  
Job time: 68 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:04 ; Search time 45.68 Seconds  
(without alignments)  
539.578 Million cell updates/sec

Title: US-09-407-430-2

Perfect score: 2058  
Sequence: 1 ETHVTVGSAGHTVSGFVSL.....ADARVCSCLWMLLSQAEA 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_66:\*  
1: piri:\*  
2: piri:\*  
3: piri:\*  
4: piri:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058	100.0	3011	1 GNMVVC3	genome polyprotein
2	1943	94.4	3011	1 S40770	genome polyprotein
3	1926	93.6	3011	1 GNMVCH	genome polyprotein
4	1817	88.3	716	2 JQ1366	polyprotein - hepa
5	1717	83.4	3010	1 GNMVTC	genome polyprotein
6	1710	83.1	3010	1 GNMVTC	genome polyprotein
7	1706	82.9	3010	1 GNMVTC	genome polyprotein
8	1696	82.4	782	2 S19875	genome polyprotein
9	1687	82.0	782	2 S19876	genome polyprotein
10	1674	81.3	782	2 S18032	genome polyprotein
11	1673	81.3	3010	1 S18030	genome polyprotein
12	1672	81.2	782	2 S18031	genome polyprotein
13	1671	81.2	3010	1 A45573	genome polyprotein
14	1651.5	80.2	3014	1 JC5620	genome polyprotein
15	1641	79.7	787	2 PN0677	hypothetical prote
16	1555	75.6	3033	1 GNMVJ8	genome polyprotein
17	1531	74.4	3033	1 JQ1303	genome polyprotein
18	1458	70.8	350	2 S35631	genome polyprotein
19	1325	64.4	640	2 JQ1584	genome polyprotein
20	1001	48.6	234	2 S32742	genome polyprotein
21	966	46.9	235	2 S32747	genome polyprotein
22	942	45.8	237	2 S32744	genome polyprotein
23	806	39.2	415	2 PC4402	genome polyprotein
24	689	33.5	513	2 A44150	envelope protein - structural protein
25	688.5	33.5	876	2 PC2219	polyptide - hepa
26	623	30.3	138	2 S24080	envelope protein -
27	609	29.6	138	2 S24081	envelope protein -
28	572	27.8	138	2 S24074	envelope protein -
29	568	27.6	138	2 S24075	envelope protein -

## ALIGNMENTS

RESULT 1

GNMVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 26-May-2000

C:Accession: A39166; PQ0403; PQ0404

R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A:Title: Genetic organization and diversity of the hepatitis C virus.

A:Reference number: A39166; MUID:91172826

A:Accession: A39166

A:Molecule type: mRNA

A:Residues: 1-3011 <CHO>

A:Cross-references: GB:M62321; NID:g329873; PTDN:AAA45676.1; PTD:g329874

R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yip,

J. Gen. Virol. 73, 1131-1141, 1992

A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t

A:Reference number: PQ0393; MUID:92268871

A:Accession: PQ0403

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CHA>

A:Cross-references: DBJ:DI0128

A:Experimental source: isolates E-b16

A:Accession: PQ0404

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1577-1633 <CH2>

A:Experimental source: isolates E-b17

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru

F:116-191/Product: capsid protein C #status predicted <CPC>

F:192-389/Product: major envelope protein M #status predicted <EPM>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: hepacivirin #status predicted <NS3>

F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>

F:1312-1317/Region: nucleotide-binding motif A (P-loop)

F:1316-1319/Region: DXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077

Query Match

Best Local Similarity 100.0%; Score 2058; DB 1; Length 3011;

Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ETHVTGSGAGHTVSGFVSLIAPGAKQNVQLINTNGSWHLNSTALNCNDSINTGWLGLFY	60
Db	384	ETHVTGSGAGHTVSGFVSLIAPGAKQNVQLINTNGSWHLNSTALNCNDSINTGWLGLFY	443
Qy	61	HHKFNSGCGPERLASCRPLTDFDQGWGPISYANGSGPDORPCYCHWYPPKCGIVPAKSV	120
Db	444	HHKFNSGCGPERLASCRPLTDFDQGWGPISYANGSGPDORPCYCHWYPPKCGIVPAKSV	503
Qy	121	GPVYCFPTSPVWYGTTRDSGAPTSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV	180
Db	504	GPVYCFPTSPVWYGTTRDSGAPTSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV	563
Qy	181	CGAPPCVIGGAGNNTLHCPDTCFKHPDATYSRCGSGPWITPRCLVDYDYPRLWHYPCITN	240
Db	564	CGAPPCVIGGAGNNTLHCPDTCFKHPDATYSRCGSGPWITPRCLVDYDYPRLWHYPCITN	623
Qy	241	YTIPIRMVYGGVEHRLAEACNTRGRCOLEDRDRSELSPLLLTTTQOVLPCSFPTLP	300
Db	624	YTIPIRMVYGGVEHRLAEACNTRGRCOLEDRDRSELSPLLLTTTQOVLPCSFPTLP	683
Qy	301	ALSTGLTHLHONIYDVQYLYXGVSSIASWAIKWEYVVLFLLLADARVCSCLWMMLLSQ	360
Db	684	ALSTGLTHLHONIYDVQYLYXGVSSIASWAIKWEYVVLFLLLADARVCSCLWMMLLSQ	743
Qy	361	AEA 363	
Db	744	AEA 746	

RESULT 2  
 genome polyprotein - hepatitis C virus  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 16-Jun-2000  
 C:Accession: S40770; PC1285  
 R:Okamoto, H.  
 submitted to the EMBL Data Library, March 1992  
 A:Reference number: S40770  
 A:Accession: S40770  
 A:Molecule type: genomic RNA  
 A:Residues: 1-3011 <OKA>  
 A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587  
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda  
 Jpn. J. Exp. Med. 60, 167-177, 1990  
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.  
 A:Reference number: PC1284; MUID:91013116  
 A:Accession: PC1285  
 A:Molecule type: genomic RNA  
 A:Residues: 1-513 <OK2>  
 A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512  
 A:Experimental source: isolate HC-J1  
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; trans  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <MEE>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
 F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DRXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match	94.4%;	Score 1943;	DB 1;	Length 3011;
Best Local Similarity	93.4%;	Pred. No. 4.7e-146;		
Matches 339: Conservative	11;	Mismatches 13;	Indels	0;
Gaps	0;			

Qy	1	ETHVTGSGAGHTVSGFVSLLAPGAKQNVLINTNGSHLNLSTALNCNDSLNTGWLGLAFY	60
Dd	384	ETIVSGGOAARASGLYSLFPPGAKQNIQLINTNGSHINSTALNCNESLNTGWLAGLIY	443
Qy	61	HHKFNSSGCPERLASCRPLTDFDQGWCPISYANGSGPDQRPCYCHWYPKPGCIGVIPAKSVC	120
Dd	444	QHKNSSGCPERLASCRRLTDFDQCWGCPISHANGSGPDQRPCYCHWYPKPGCIGVIPAKSVC	503
Qy	121	GPVYCFTPSPVVVGTTDRSGAPTYSWGENDTDVFVLNNTRRPLLGNWFGCTWMNSTGFTKV	180
Dd	504	GPVYCFTPSPVVVGTTDRSGAPTYNWGANDTDVFLNNTRRPLLGNWFGCTWMNSTGFTKV	563
Qy	181	CGAPPCVIGGAGNWTLCPTDCFRKHFDATYSRCGSGPWITPRCLVDVPYRLWHYPCTIN	240
Dd	564	CGAPPCVIGGGNNTLCPTDCFRKHFEATYSRCGSGPWITPRCLVDVPYRLWHYPCTIN	623
Qy	241	YTIFKIRMYGVGEHRLAEACNWTGRGERCDLEDRSELSPLLLITTTQMOWLPSCFTTL	300
Dd	624	YTIFKVRMYGVGEHRLDAACNWTGRGERCDLEDRSELSPLLLLSTTTQMOWLPSCFTTL	683
Qy	301	ALSTGLIHLHONIYDVQYLGVGSSIASWAIKVEYVLLFLLADARVCSLWNMLLIQS	360
Dd	684	ALSTGLIHLHONIYDVQYLGVGSSIASWAIKWEYVLLFLLADARVCSLWNMLLIQS	743
Qy	361	AFA 363	
Dd	744	AFA 746	
 RESULT 3 GNWVCH			
genome polyprotein - hepatitis C virus (strain H)			
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstruc			
tural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
A:Note: host Homo sapiens (man)			
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 26-May-2000			
C:Accession: A36814; A41546			
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.			
submitted to GenBank, July 1992			
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus			
A:Reference number: A36814			
A:Accession: A36814			
A:Molecule type: genomic RNA			
A:Residues: 1-3011 <NC>			
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738			
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.			
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991			
A>Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp			
A:Reference number: A41546; MUID:92052256			

A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
C:Keywords: Amp; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct  
F:1\_115/Product: capsid protein C #status predicted <CPC>  
F:1\_115/Product: capsid protein M #status predicted <EPM>  
F:116-191/Product: envelope protein E #status predicted <WEE>  
F:192-389/Product: major envelope protein E #status predicted <N1>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitisvirin #status predicted <NS3>  
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:1614-2011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196-209/234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 121  
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 121

Query Match 93.6%; Score 1926; DB 1; Length 3011;  
Best Local Similarity 93.1%; Pred. No. 1.le-144;



Matches 338; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTIVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFY 60  
Db 384 ETHVTGAGNAGTTAGVGLLTPGAKQNIQIINTNGSWHLNSTALNCNDSLNTGWLAGLFY 443

QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIVPAKSV 120  
Db 444 OHKFNSSGCGPERLASCRRLTDFAGWGPISYANGSGLDERPYCHWYPPRPGIVPAKSV 503

QY 121 GPVYCFPTSPVVVGTDRSGAPTYSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 504 GPVYCFPTSPVVVGTDRSGAPTYSWGANDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 563

QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWHYPCTIN 240  
Db 564 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWHYPCTIN 623

QY 241 YTIKIRMYVGGVEHRLAEACNWTGRGCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300  
Db 624 YTIKIRMYVGGVEHRLAEACNWTGRGCDLDRSELSPLLLTTTQWQVLPSCFTTLP 683

QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 360  
Db 684 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 743

QY 361 AEA 363  
Db 744 AEA 746

RESULT 4

QJ1366  
polyprotein - hepatitis C virus (French isolate) (fragments)  
C:Species: hepatitis C virus  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 29-May-1998  
C:Accession: JQ1366  
R:Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.  
J. Gen. Virol. 72, 2557-2561, 1991  
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication  
A:Reference number: JQ1366; MUID:92013977  
A:Accession: JQ1366  
A:Molecule type: genomic RNA  
A:Residues: 1-716 <KRE>  
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
C:Keywords: glycoprotein; polyprotein  
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 88.3%; Score 1817; DB 2; Length 716;  
Best Local Similarity 86.7%; Pred. No. 9.9e-137;  
Matches 306; Conservative 30; Mismatches 17; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTIVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFY 60  
Db 51 ETVTTGGSTARTTQGLVSLFSGAKQDIQIINTNGSWHLNSTALNCNDSLNTGWLAGLFY 110

QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIVPAKSV 120  
Db 111 YHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGTGPYCHWYPPKPGIVPAQTV 170

QY 121 GPVYCFPTSPVVVGTDRSGAPTYSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 171 GPVYCFPTSPVVVGTDRSGAPTYSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 230

QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWHYPCTIN 240  
Db 231 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWHYPCTIN 290

QY 241 YTIKIRMYVGGVEHRLAEACNWTGRGCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300  
Db 291 YTIKIRMYVGGVEHRLAEACNWTGRGCDLDRSELSPLLLTTTQWQVLPSCFTTLP 350

QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMW 353  
Db 351 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMW 403

RESULT 5  
GNWTC

genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstru  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 26-May-2000  
C:Accession: A38465  
R:Takamiyawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E  
J. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from hum  
A:Reference number: A38465; MUID:91140698  
A:Accession: A38465

A:Molecule type: genomic RNA  
A:Residues: 1-3010 <TAK>  
A:Cross-references: EMBL:N58335; NID:g329770; PIDN:AAA72945.1; PID:g329771  
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEP>  
F:390-1006/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077

Query Match 83.4%; Score 1717; DB 1; Length 3010;  
Best Local Similarity 80.4%; Pred. No. 4.2e-128;  
Matches 292; Conservative 29; Mismatches 42; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTIVSGFVSLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFY 60  
Db 384 ETHVTGGAQAKTNNRLVSMFASGSPQKIQIINTNGSWHLNSTALNCNDSLOTGFLAALFY 443

QY 61 HHKFNSSGCGPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIVPAKSV 120  
Db 444 THSFNSSGCGPERMAOQRTIDKFDQGWGPIYAESRSDDQRPYCHWYPPQCTIVPASEVC 503

QY 121 GPVYCFPTSPVVVGTDRSGAPTYSWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 504 GPVYCFPTSPVVVGTDRFGVPTYRWGENETDVLNNTRPPOGNWFGCTWMNSTGFTKT 563

QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSRGSGPWTTPRCLVDYPRYLWHYPCTIN 240  
Db 564 CGGPNIGGVGNNTLTCTDCFRKHPEATYTKCGSGPMLTPRCMVDPYRLWHYPCTVN 623

QY 241 YTIKIRMYVGGVEHRLAEACNWTGRGCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300  
Db 624 YTIKIRMYVGGVEHRLAEACNWTGRGCDLDRPELSPLLSTTQWQVLPSCFTTLP 683

QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWLLISQ 360  
Db 684 ALSTGLIHLHQNIVDVQVLYGIGSAVSAIKWEYVVLFLLLADARVCSCLMWLLISQ 743

QY 361 AEA 363  
Db 744 AEA 746

RESULT 6

## GNMWTV

genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)  
C:Keywords: ATP; capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: major envelope protein M #status predicted <MEE>  
F:192-389/Product: major envelope protein M #status predicted <NS2>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS2>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS3>  
F:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the  
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 207

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 26-May-2000

C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.

Virolgy 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the

A:Reference number: A40244; MUID:92230206

A:Accession: A40244

A:Molecule type: genomic RNA

A:Residues: 1-3010 <CH>

A:Cross-references: GB:M84754

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: major envelope protein M #status predicted <MEE>

F:192-389/Product: major envelope protein M #status predicted <NS2>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS2>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS3>

F:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the

F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>

F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 207

Query Match 83.18; Score 1710; DB 1; Length 3010;

Best Local Similarity 79.6%; Pred. No. 1.5e-127;

Matches 288; Conservative 33; Mismatches 41; Indels 0; Gaps 0;

Qy 2 THVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFYH 61

Db 385 TIVSGTIVARTHSLSLFTGASQKIQLVNTNGSWHLNRTALNCNDSLTGWLGLFYH 444

Qy 62 HFNSSGCPERLASCRPLTDFDQGWGPIYANGSGDPQRPYCWHPKPGCIVPAKSVCG 121

Db 445 HRFNAGCPCERMAASCRPIDQGWGPIYANGSGDPQRPYCWHPKPGCIVPAKSVCG 504

Qy 122 PVYCTPSPVVGTTDRSGAPTSWGENDTDFVFLNTRPPLGNWFGCTWNSGTFTKVC 181

Db 505 PVYCTPSPVVGTTDRSGAPTSWGENDTDFVFLNTRPPLGNWFGCTWNSGTFTKVC 564

Qy 182 GAPPVIGGAGNNTLHCPDTCFRKHDPATYSCGSGPWITPRCLVDYPIRLWHPCTINY 241

Db 565 GPPCNIGGGNNTLVCPDTCFRKHDPATYSCGSGPWITPRCLVDYPIRLWHPCTINY 624

Qy 242 TIFKIRMYGVGVEHRLAECNWTGERCDLDRDRSELSPLLLTTQWQVLPCTFTLPA 301

Db 625 TIFKIRMYGVGVEHRLAECNWTGERCDLDRDRSELSPLLLTTQWQVLPCTFTLPA 684

Qy 302 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 361

Db 685 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 744

Qy 362 EA 363

Db 745 EA 746

RESULT 7

GNMVCJ

genome polyprotein - hepatitis C virus (strain J)

N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural

protein NS4a; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jun-2000

C:Accession: A39253; PS0086

R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patient

A:Reference number: A39253; MUID:91088550

A:Accession: A39253

A:Molecule type: genomic RNA

A:Residues: 1-3010 <KAT>

A:Cross-references: GB:D90208; NID:G221610; PIDN:BAA14233.1; PID:G221611

R:Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 65B, 219-223, 1989

A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v

A:Reference number: PS0085

A:Accession: PS0086

A:Molecule type: genomic RNA

A:Residues: 2650-2707 <KAT>

A:Experimental source: Japanese isolate

C:Comment: The cleavage sites of this polyprotein have not been determined.

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: major envelope protein M #status predicted <MEE>

F:192-389/Product: major envelope protein E #status predicted <NS2>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS2>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS3>

F:1007-1615/Product: hepatitis C virus genome: sequence determination and mapping the

F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>

F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 224

Query Match 82.9%; Score 1706; DB 1; Length 3010;

Best Local Similarity 78.7%; Pred. No. 3.1e-127;

Matches 285; Conservative 39; Mismatches 38; Indels 0; Gaps 0;

Qy 2 THVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFYH 61

Db 385 THVTGGRVASSTQSLVSWLSQPSQKIQLVNTNGSWHLNRTALNCNDSLTGWLGLFYH 444

Qy 62 HFNSSGCPERLASCRPLTDFDQGWGPIYANGSGDPQRPYCWHPKPGCIVPAKSVCG 121

Db 445 HRFNAGCPCERMAASCRPIDQGWGPIYANGSGDPQRPYCWHPKPGCIVPAKSVCG 504

Qy 122 PVYCTPSPVVGTTDRSGAPTSWGENDTDFVFLNTRPPLGNWFGCTWNSGTFTKVC 181

Db 505 PVYCTPSPVVGTTDRSGAPTSWGENDTDFVFLNTRPPLGNWFGCTWNSGTFTKVC 564

Qy 182 GAPPVIGGAGNNTLHCPDTCFRKHDPATYSCGSGPWITPRCLVDYPIRLWHPCTINY 241

Db 565 GPPCNIGGGNNTLVCPDTCFRKHDPATYSCGSGPWITPRCLVDYPIRLWHPCTINY 624

Qy 242 TIFKIRMYGVGVEHRLAECNWTGERCDLDRDRSELSPLLLTTQWQVLPCTFTLPA 301

Db 625 TIFKIRMYGVGVEHRLAECNWTGERCDLDRDRSELSPLLLTTQWQVLPCTFTLPA 684

Qy 302 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 361

Db 685 LSTGLIHLHQNIVDQYLYGVGSSSTASWAIKWEYVLLFLLLADARVCSCLWMLLSQA 744

Qy 362 EA 363

Db 745 EA 746

RESULT 8

S19875

genome polyprotein - hepatitis C virus (isolate JK3) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein  
 C:Species: hepatitis C virus  
 A:Variety: isolate JK3  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 26-Aug-1999  
 C:Accession: S19875  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso  
 A:Reference number: S18029

A:Accession: S19875

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61592; NID:g59482; PIDN:CAA43789.1; PID:g59483

A:Experimental source: isolate JK3

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 82.4%; Score 1696; DB 2; Length 782;  
 Best Local Similarity 78.2%; Pred. No. 4.5e-127;  
 Matches 284; Conservative 38; Mismatches 41; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 60

Db 384 QTRVTAQVGRITSSLTSTFTPGPSONIQLVNSGSHINRTALSCNDSLKTGFLAALFY 443

QY 61 HHKFNSSGCGPERLASRPLTDFDQGWGPIYSANGSGDPQRPYCHWYPPKPCGIVPAKSYVC 120

Db 444 THKFNASGCGPERMASCRSIDTDFDQGWGPIHVPNTDQKPYCHWYAPRPGCGIVPASQVC 503

QY 121 GPVYCTPSPVVVGTTRDSGAPYSNGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 180

Db 504 GPVYCTPSPVVVGTTRDSGAPYTWGENETDVLNTRPPLGNMFGCTWMNSTGFTKT 563

QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSCGSGPWITPRCLVDYPRVRLWHYPCTIN 240

Db 564 CGGPPCNIGGAGNTLHCPTDCFRKHPDATYSCGSGPWITPRCLVDYPRVRLWHYPCTIN 623

QY 241 YTIKIRMYVGVGVEHLEAACNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300

Db 624 FSVFKIRMYVGVGVEHLEAACNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 683

QY 301 ALSTGLIHLHONIVDQVLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360

Db 684 ALSTGLIHLHONIVDQVLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 743

QY 361 AEA 363

Db 744 AEA 746

RESULT 9

S19876

genome polyprotein - hepatitis C virus (isolate JK5) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK5

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 26-Aug-1999

C:Accession: S19876

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso

A:Reference number: S18029

A:Accession: S19876

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61595; NID:g59486; PIDN:CAA43792.1; PID:g59487

A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural  
 F:1-191/Product: core protein #status predicted <MAT1>  
 F:192-383/Product: envelope protein 1 #status predicted <MAT2>  
 F:384-733/Product: NS1/E2 protein #status predicted <MAT3>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 82.0%; Score 1687; DB 2; Length 782;  
 Best Local Similarity 78.8%; Pred. No. 2.3e-126;  
 Matches 286; Conservative 34; Mismatches 43; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 60

Db 384 QTRVTAQVGRITSSLTSTFTPGPSONIQLVNSGSHINRTALSCNDSLKTGFLAALFY 443

QY 61 HHKFNSSGCGPERLASRPLTDFDQGWGPIYSANGSGDPQRPYCHWYPPKPCGIVPAKSYVC 120

Db 444 THKFNASGCGPERMASCRSIDTDFDQGWGPIHVPNTDQKPYCHWYAPRPGCGIVPASQVC 503

QY 121 GPVYCTPSPVVVGTTRDSGAPYSNGENDTDVFLNTRPPLGNMFGCTWMNSTGFTKV 180

Db 504 GPVYCTPSPVVVGTTRDSGAPYTWGENETDVLNTRPPLGNMFGCTWMNSTGFTKT 563

QY 181 CGAPPCVIGGAGNTLHCPTDCFRKHPDATYSCGSGPWITPRCLVDYPRVRLWHYPCTIN 240

Db 564 CGGPPCNIGGAGNTLHCPTDCFRKHPDATYSCGSGPWITPRCLVDYPRVRLWHYPCTIN 623

QY 241 YTIKIRMYVGVGVEHLEAACNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 300

Db 624 FTFIKIRMYVGVGVEHLEAACNWTGERCDLDRSELSPLLLTTTQWQVLPSCFTTLP 683

QY 301 ALSTGLIHLHONIVDQVLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360

Db 684 ALSTGLIHLHONIVDQVLYGVGSSSTASNAIKWEYVVLFLLLADARVCSCLMMMLLSQ 743

QY 361 AEA 363

Db 744 AEA 746

RESULT 10

S18032

genome polyprotein - hepatitis C virus (isolate JK4) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK4

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 29-May-1998

C:Accession: S18032

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus

A:Reference number: S18029

A:Accession: S18032

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:Cross-references: EMBL:X61594

A:Experimental source: isolate JK4

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: envelope protein 1 #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 81.3%; Score 1674; DB 2; Length 782;  
 Best Local Similarity 78.7%; Pred. No. 2.5e-125;  
 Matches 285; Conservative 33; Mismatches 44; Indels 0; Gaps 0;

QY 2 THVTGSGAGTYSVGFVSLAPCAKONVOLINTNGSHLNSTALNCNDSLNTGWLGLFY 61

Db 385 TTVSGGHASQITRGVTSPFSPGSAQKIQLVNTNGSHINRTALNCNDSINTGFFAALFYA 444



QY 182 GAPPVIGGAGNNTLHCPDTCFRKHDPATYRCGSGPWITPRCLVDYDYPRLWHYPCITNY 241  
 Db 565 EGPCNIGGVGNNTLTCTDTCFRKHDPATYTKCGSGPMLTPRCMVHYPYRLWHYPCITVNF 624

QY 242 TIFKIRMYGVGVEHRLAECANWTRGERCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLPA 301  
 Db 625 TIFKIRMYGVGVEHRLAECANWTRGERCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLPA 684

QY 302 LSTGLIHLHONIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCSCIMMMLISQA 361  
 Db 685 LSTGLIHLHONIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCACIMMMLISQA 744

QY 362 EA 363  
 Db 745 EA 746

RESULT 13  
 A45573  
 genome polyprotein - hepatitis C virus (strain JT)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS5 (nonstructural protein NS4b; nonstructural protein NS4b; nonstructural protein NS5)  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 16-Jun-2000  
 C:Accession: A45573  
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Virus Res. 23, 39-53, 1992  
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  
 A:Reference number: A45573; MUID:92295714  
 A:Accession: A45573  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3010 <N>  
 A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PID:BAA01943.1; PID:g221613  
 A:Experimental source: HCV-JT  
 A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:106207)  
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
 F:115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
 F:1230-1493/Product: DEAD/H box helicase homology <DEAD>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1317/Region: nucleotide-binding motif B  
 F:1316-1319/Region: DEXH motif  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 81.2%; Score 1671; DB 1; Length 3010;  
 Best Local Similarity 78.5%; Pred. No. 1.9e-124;  
 Matches 284; Conservative 29; Mismatches 49; Indels 0; Gaps 0;

QY 2 THVTGSGAGHTVGVFSLAPAGKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 61  
 Db 385 THVTGSGAGHTVGVFSLAPAGKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 444

QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCYWHYPPKCGIVPAKSVG 121  
 Db 445 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCYWHYPPKCGIVPAKSVG 504

QY 122 PVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLGNWFGCTWMNSTGFTKVC 181  
 Db 505 PVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLGNWFGCTWMNSTGFTKVC 564

QY 182 GAPPVIGGAGNNTLHCPDTCFRKHDPATYRCGSGPWITPRCLVDYDYPRLWHYPCITNY 241  
 Db 565 EGPCNIGGVGNNTLTCTDTCFRKHDPATYTKCGSGPMLTPRCMVHYPYRLWHYPCITVNF 624

QY 242 TIFKIRMYGVGVEHRLAECANWTRGERCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLPA 301  
 Db 625 TIFKIRMYGVGVEHRLAECANWTRGERCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLPA 684

QY 302 LSTGLIHLHONIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCSCIMMMLISQA 361  
 Db 685 LSTGLIHLHONIVDVQVLYGVGSSIASWAIKWEYVLLFLLLDARVCACIMMMLISQA 744

QY 362 EA 363  
 Db 745 EA 746

RESULT 14  
 JC5620  
 genome polyprotein - hepatitis C virus (isolate EUH1480)  
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein NS5 (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
 C:Species: hepatitis C virus  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
 C:Accession: JC5620  
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M. Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A:Title: The complete coding sequence of hepatitis C virus genome polyprotein; serine proteinase; tra  
 A:Reference number: JC5620; MUID:97366593  
 A:Accession: JC5620  
 A:Molecule type: mRNA  
 A:Residues: 1-3014 <CHA>  
 A:Cross-references: GB:Y13184  
 A:Experimental source: genotype 5a, which predominates in South Africa  
 A:Note: the translation of the nucleotide sequence is not complete in this paper  
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
 C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EPM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:384-408/Region: hypervariable #status predicted  
 F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1008-1616/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
 F:1231-1494/Domain: DEAD/H box helicase homology <DEAD>  
 F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
 F:1313-1318/Region: nucleotide-binding motif B  
 F:1317-1320/Region: DEXH motif  
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4>  
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 80.2%; Score 1651.5; DB 1; Length 3014;  
 Best Local Similarity 75.8%; Pred. No. 6.7e-123;  
 Matches 275; Conservative 39; Mismatches 48; Indels 1; Gaps 1;

QY 2 THVTGSGAGHTVGVFSLAPAGKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 61  
 Db 385 THVTGSGAGHTVGVFSLAPAGKQVQLINTNGSHWLNSTALNCNDSLNTGLAGLYFH 444

QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCYWHYPPKCGIVPAKSVG 120  
 Db 445 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGP-DORPCYWHYPPKCGIVPAKSVG 504

QY 121 GPVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLGNWFGCTWMNSTGFTKVC 180  
 Db 505 GPVYCFPTSPVVGTTDRSGAPYSWENDTDVFLNTRPPLGNWFGCTWMNSTGFTKVC 564

QY 181 GAPPVIGGAGNNTLHCPDTCFRKHDPATYRCGSGPWITPRCLVDYDYPRLWHYPCITNY 240  
 Db 565 GAPPVIGGAGNNTLHCPDTCFRKHDPATYRCGSGPWITPRCLVDYDYPRLWHYPCITNY 624

QY 241 YTIKIRMYGVGVEHRLAECANWTRGERCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLPA 300  
 Db 565 YTIKIRMYGVGVEHRLAECANWTRGERCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLPA 624

Job time: 152 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:33 ; Search time 25.34 Seconds  
(without alignments)  
462.619 Million cell updates/sec

Title: US-09-407-430-2

Perfect score: 2058

Sequence: 1 ETHVTGSGAGHTVSGFVSL.....ADARVCSLWMLISQAEA 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2058	100.0	3011	1	POLG_HCV1
2	1926	93.6	3011	1	POLG_HCVH
3	1717	83.4	3010	1	POLG_HCVBK
4	1710	83.1	3010	1	POLG_HCVTW
5	1706	82.9	3010	1	POLG_HCVJA
6	1671	81.2	3010	1	POLG_HCVJT
7	1555	75.6	3033	1	POLG_HCVJB
8	1531	74.4	3033	1	POLG_HCVJB
9	1508	73.3	737	1	POLG_HCVJ7
10	1473	71.6	737	1	POLG_HCVJ5
11	555	27.0	520	1	POLG_HCVH4
12	553	26.9	513	1	POLG_HCVJ2
13	548	26.6	520	1	POLG_HCVHK
14	238	11.6	321	1	POLG_HCVH8
15	225	10.9	321	1	POLG_HCVH7
16	216	10.5	138	1	POLG_HCVH0
17	199	9.7	309	1	POLG_HCVH7
18	110.5	5.4	1700	1	BAR3_CHITE
19	107.5	5.2	1046	1	PSTA_DICDI
20	103	5.0	1172	1	TSP2_MOUSE
21	100	4.9	640	1	UROM_HUMAN
22	97	4.7	752	1	C02_HUMAN
23	95.5	4.6	661	1	F13B_HUMAN
24	95	4.6	577	1	VGLE_PPRV1
25	94	4.6	555	1	DP87_DICDI
26	94	4.6	1172	1	TSP2_HUMAN
27	93	4.5	551	1	AMVB_THETU
28	92	4.5	592	1	INV2_DAUCA
29	92	4.5	1077	1	SM5A_MOUSE
30	91.5	4.4	571	1	HEMA_P12H
31	91.5	4.4	571	1	HEMA_P12H
32	91	4.4	718	1	CDGT_BACSS
33	91	4.4	2871	1	FBN1_HUMAN

34 90.5 4.4 784 1 YAV2\_XANCV  
35 90 4.4 1170 1 TSP2\_BOVIN  
36 90 4.4 2871 1 FBNI\_MOUSE  
37 89 4.3 825 1 BISC\_HAFIN  
38 88.5 4.3 2182 1 POLG\_CXB1J  
39 87 4.2 532 1 AD50\_BOVIN  
40 87 4.2 1178 1 TSPI\_CHICK  
41 87 4.2 2871 1 FBNI\_BOVIN  
42 86.5 4.2 495 1 MLP2\_DROME  
43 86 4.2 1093 1 SM5B\_MOUSE  
44 84.5 4.1 643 1 UROM\_BOVIN  
45 84 4.1 1148 1 VGLM\_PUUMH

#### ALIGNMENTS

RESULT 1

POLG\_HCV1

ID POLG\_HCV1 STANDARD; PRT: 3011 AA.

AC P26664;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);

DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2

DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)

DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);

DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);

DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)

DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);

OS Hepatitis C virus (isolate 1) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-91172826; PubMed-1848704;

RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,

RA Gallegos C., Colt D., Medina-Selby A., Barr P.J., Weiner A.J.,

RA Bradley D.W., Kuo G., Houghton M.;

RT "Genetic organization and diversity of the hepatitis C virus.";

CC Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M62321; AAA45676.1; -

CC PIR; A39166; GNVVC3.

CC HSSP; P27958; 1HEI.

CC MEROPS; S29.001; -

CC MEROPS; U39.001; -

CC INTERPRO; IPR000745; -

CC INTERPRO; IPR001490; -

CC INTERPRO; IPR002166; -

CC INTERPRO; IPR002518; -

CC INTERPRO; IPR002519; -

CC INTERPRO; IPR002521; -

CC INTERPRO; IPR002522; -

CC INTERPRO; IPR002531; -

CC INTERPRO; IPR002868; -

DR PFAM; PF01560; HCV\_NS1; 1.  
 DR PFAM; PF01538; HCV\_NS2; 1.  
 DR PFAM; PF01006; HCV\_NS4a; 1.  
 DR PFAM; PF01001; HCV\_NS4b; 1.  
 DR PFAM; PF01506; HCV\_NS5a; 1.  
 DR PFAM; PF00998; HCV\_RdRP; 1.  
 DR PFAM; PF01543; HCV\_capsid; 1.  
 DR PFAM; PF01542; HCV\_core; 1.  
 DR PFAM; PF01539; HCV\_env; 1.  
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.0%; Score 2058; DB 1; Length 3011;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-167;  
 Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVGTSGAGTVSGVSVLLAPGAKONVGLINTNGSWHLNSTALNCNDSLNTGLWLAGLFY 60  
 DB 384 ETHVGTSGAGTVSGVSVLLAPGAKONVGLINTNGSWHLNSTALNCNDSLNTGLWLAGLFY 443  
 QY 61 HHKFNSSGGPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCHWYPPKPGCIVPAKSYVC 120  
 DB 444 HHKFNSSGGPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCHWYPPKPGCIVPAKSYVC 503  
 QY 121 GPVYCTFPSPVVVGTDRSGATYSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 180  
 DB 504 GPVYCTFPSPVVVGTDRSGATYSYSGENDTDVFLNNTRPPLGNWFGCTWNNSTGFTKV 563  
 QY 181 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGPWITPRCLVDYDYPRLWHYPCPTIN 240  
 DB 564 CGAPPCVIGAGNNTLHCPTDCFRKHPDATYRCGSGPWITPRCLVDYDYPRLWHYPCPTIN 623  
 QY 241 YTIKIRMYVGVGVEHLEACNWTGRCERCDLEDRDRSELSPILLITTTQWQLPCSFITLP 300

DB 624 YTIKIRMYVGVGVEHLEACNWTGRCERCDLEDRDRSELSPILLITTTQWQLPCSFITLP 683  
 QY 301 ALSTGLIHQHONIVDYQYLYGVSSSTASWAIKWEYVYVLLFLLADARVCSCLMMMLLSQ 360  
 DB 684 ALSTGLIHQHONIVDYQYLYGVSSSTASWAIKWEYVYVLLFLLADARVCSCLMMMLLSQ 743  
 QY 361 AEA 363  
 DB 744 AEA 746  
 RESULT 2  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 ID POLG\_HCVH  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)  
 DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE-92052256; PubMed-1659800;  
 RX Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RA "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 [2]  
 RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RP MEDLINE-9731322; PubMed-9187654;  
 RX Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RA "Structure of the hepatitis C virus RNA helicase domain.";  
 RT Nat. Struct. Biol. 4:463-467(1997).  
 [3]  
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RP MEDLINE-98154321; PubMed-9493270;  
 RX Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RA "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.  
 CC -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.  
 CC -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY C19.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. usage by and for commercial





RT virus: a 2.2-A resolution structure in a hexagonal crystal form.";  
 RL Protein Sci 7:837-847(1998).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M58335; AAA72945.1; -  
 PIR: A38465; GNWVTC.  
 PDB: 1A1Q; 25-MAR-98.  
 PDB: 1JXP; 14-JAN-98.  
 PDB: 1NS3; 08-APR-98.  
 MEROPS: S29.001; -  
 MEROPS: U39.001; -  
 INTERPRO: IPRO00745; -  
 INTERPRO: IPRO01490; -  
 INTERPRO: IPRO02166; -  
 INTERPRO: IPRO02518; -  
 INTERPRO: IPRO02519; -  
 INTERPRO: IPRO02521; -  
 INTERPRO: IPRO02522; -  
 INTERPRO: IPRO02531; -  
 INTERPRO: IPRO02868; -  
 PFAM: PF01560; HCV\_NS1; 1.  
 PFAM: PF01538; HCV\_NS2; 1.  
 PFAM: PF01006; HCV\_NS4a; 1.  
 PFAM: PF01001; HCV\_NS4b; 1.  
 PFAM: PF01506; HCV\_NS5a; 1.  
 PFAM: PF00998; HCV\_RdrP; 1.  
 PFAM: PF01543; HCV\_capsid; 1.  
 PFAM: PF01542; HCV\_core; 1.  
 PFAM: PF01539; HCV\_env; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;  
 Query Match 83.4%; Score 1717; DB 1; Length 3010;  
 Best Local Similarity 80.4%; Pred. No. 6e-138;  
 Matches 292; Conservative 29; Mismatches 42; Indels 0; Gaps 0;  
 QY 1 ETHVTGSGAGTSGVFYSLAPGAKQNVQLINTNGSHLNTALNCNDSLTGWLGLFY 60  
 DB 384 DTHVTGGAQAQAKTTRNLVSMFASGSPQKIQLINTNGSHINRTALNCNDSLTGFLAALFY 443  
 QY 61 HHKFNSSGCPERLASRPLTDFOGWGPISYANGSGDPQRPYCHYPPKPGIVPAKSVC 120  
 DB 444 THSFNSSGCPERMAQCRITDKFDGWCGRPYTAESSRSDQRPYCHYPPPOCTIVPASEVC 503  
 QY 121 GPVYCFPTSPVVGTTDRSGAPTYSWGNEDTVFLNNTTRPLGNWFCCTWMNSTGFTKV 180  
 DB 504 GPVYCFPTSPVVGTTDRSGAPTYSWGNEDTVFLNNTTRPLGNWFCCTWMNSTGFTKT 563  
 QY 181 CGAPPVIGGAGNNTLHCPTDCFRKHPDATYSRGGSGPWITPCLVDYPRVLRHYPCTIN 240  
 DB 564 CGPPCNIGVGNNLTCTPTDCFRKHPDATYSRGGSGPWITPCLVDYPRVLRHYPCTIN 623  
 QY 241 YTIKIRMYGVGVEHRLAEACNWTNRGERCDLEDRDELSPILLITTTQWQVLPSCFTLP 300  
 DB 624 FTIFKVRMYGVGVEHRLAEACNWTNRGERCDLEDRDELSPILLITTTQWQVLPSCFTLP 683  
 QY 301 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVLLFLLADARVCSLWMLLIQS 360  
 DB 684 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVLLFLLADARVCSLWMLLIQS 743  
 QY 361 AEA 363  
 DB 744 AEA 746

RESULT 4  
 POLG\_HCVTW STANDARD; PRT; 3010 AA.  
 AC P29846;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)  
 DE (EC 3.4.22.-); PROTEINASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate Taiwan) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92230206; PubMed-1314449;  
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;  
 RA "The Taiwanese hepatitis C virus genome: sequence determination and  
 RA mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RL Virology 188:102-113(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND MRNA.  
-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: M84754; -; NOT\_ANNOTATED\_CDS.  
PIR: A40244; GNWVTV.  
HSP: P27958; 1HEI.  
MEROPS: S29.001; -.  
MEROPS: U39.001; -.  
INTERPRO: IPR000745; -.  
INTERPRO: IPR001490; -.  
INTERPRO: IPR002166; -.  
INTERPRO: IPR002518; -.  
INTERPRO: IPR002519; -.  
INTERPRO: IPR002521; -.  
INTERPRO: IPR002522; -.  
INTERPRO: IPR002531; -.  
INTERPRO: IPR002868; -.  
PFAM: PF01560; HCV\_NS1; 1.  
PFAM: PF01538; HCV\_NS2; 1.  
PFAM: PF01006; HCV\_NS4a; 1.  
PFAM: PF01001; HCV\_NS4b; 1.  
PFAM: PF01506; HCV\_NS5a; 1.  
PFAM: PF00998; HCV\_RDRP; 1.  
PFAM: PF01543; HCV\_capsid; 1.  
PFAM: PF01542; HCV\_core; 1.  
PFAM: PF01539; HCV\_env; 1.  
Polyprotein: glycoprotein; Transferase; RNA-directed RNA polymerase;  
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
Transmembrane; Nonstructural  
INIT\_MET 1  
CHAIN 1 115  
CHAIN 116 191  
CHAIN 192 383  
CHAIN 384 729  
CHAIN 730 1006  
CHAIN 1007 1615  
CHAIN 1616 1862  
CHAIN 1863 2013  
CHAIN 2014 3010  
CHAIN 3011 369  
TRANSMEM 347 369  
ACT\_SITE 1083 1083  
ACT\_SITE 1107 1107  
ACT\_SITE 1155 1155  
ACT\_SITE 1230 1237  
SITE 1316 1319  
SITE 1316 1319  
CARBOHYD 196 196  
CARBOHYD 209 209  
CARBOHYD 233 233  
CARBOHYD 234 234  
CARBOHYD 250 250  
CARBOHYD 305 305  
CARBOHYD 417 417  
CARBOHYD 423 423  
CARBOHYD 430 430  
CARBOHYD 448 448  
CARBOHYD 532 532  
CARBOHYD 540 540  
CARBOHYD 556 556  
CARBOHYD 576 576

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;  
  
Query Match 83.1%; Score 1710; DB 1; Length 3010;  
Best Local Similarity 79.6%; Pred. No. 2.4e-137;  
Matches 288; Conservative 33; Mismatches 41; Indels 0; Caps 0;  
  
QY 2 THVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFYH 61  
DB 385 TIVSGGTAVRTTSLASLFTQASQKIQIINTNGSWHLNSTALNCNDSLTGWLGLFYH 61  
QY 62 HRFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDPCYCHYPPKCGIYPAKSVCG 121  
DB 445 HRFNASGCPERMASCRSIDKFDQGWGPIITYEADIODPCYCHYAPRPGIYPAKSVCG 504  
QY 122 PVYCFTPSPVVVGTDRSGAPTYSWGENDTDFVLNTRPPPLGNWFCCTWMNSTGFTKVC 181  
DB 505 PVYCFTPSPVVVGTDRFGAPTYSWGENETDVLILNTRPPQGNWFCCTWMNSTGFTKVC 564  
QY 182 GAPPVIGGAGNNTLHCPTDCFRKHPDATYSRGGSGPWITPRCLVDYVYRLWHYPTINY 241  
DB 565 GGPPCNIGGGNNTLVCPTDCFRKHPDATYKCGSGPWLTPRCMVDYVYRLWHYPTVNF 624  
QY 242 TIFKIRMYGVGVEHRLAECNNTGRGCRDLEDRSELSPLLLTTTQWVLPSCFTLPA 301  
DB 625 TIFKIRMYGVGVEHRLAECNNTGRGCRDLEDRSELSPLLLTTTQWVLPSCFTLPA 684  
QY 302 LSTGLIHLHQNVDVQVLYGVSSIASWAIKWEYVVLFLLLADARVCSCLMMLLSOA 361  
DB 685 LSTGLIHLHQNVDVQVLYGVSSIASWAIKWEYVVLFLLLADARVCSCLMMLLSOA 744  
QY 362 EA 363  
DB 745 EA 746  
  
RESULT 5  
POLG-HCVJA STANDARD; PRT: 3010 AA.  
ID POLG-HCVJA AC P26662;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)  
DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate Japanese) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91088550; PubMed=2175903;  
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
RA Sugimura T., Shimotohno K.;  
RT "Molecular cloning of the human hepatitis C virus genome from  
RT Japanese patients with non-A, non-B hepatitis";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
RN [2]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=91192160; PubMed=1849488;  
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraliso K.,  
RA Ohkoshi S., Shimotohno K.;

RT "Molecular structure of the Japanese hepatitis C viral genome.";  
 RL FEBS Lett. 280:325-328(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D90208; BAA14233.1; -  
 CC PIR; A39253; GNWVCJ.  
 CC HSP; P27958; 1HEI.  
 CC MEROPS; S29.001; -  
 CC INTERPRO; IPR000745; -  
 CC INTERPRO; IPR001490; -  
 CC INTERPRO; IPR002166; -  
 CC INTERPRO; IPR002518; -  
 CC INTERPRO; IPR002519; -  
 CC INTERPRO; IPR002521; -  
 CC INTERPRO; IPR002522; -  
 CC INTERPRO; IPR002531; -  
 CC INTERPRO; IPR002868; -  
 CC PFAM; PF01560; HCV\_NS1; 1.  
 CC PFAM; PF01538; HCV\_NS2; 1.  
 CC PFAM; PF01006; HCV\_NS4a; 1.  
 CC PFAM; PF01001; HCV\_NS4b; 1.  
 CC PFAM; PF01506; HCV\_NS5a; 1.  
 CC PFAM; PF00998; HCV\_RGRP; 1.  
 CC PFAM; PF01543; HCV\_capsid; 1.  
 CC PFAM; PF01542; HCV\_core; 1.  
 CC PFAM; PF01539; HCV\_env; 1.  
 CC  
 CC Polyprotein; Glycoprotein; Transferrin; RNA-directed RNA polymerase;  
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 CC Transmembrane; Nonstructural  
 CC INIT\_MET 1 1  
 CC CHAIN 1 115  
 CC CHAIN 116 191  
 CC CHAIN 192 383  
 CC CHAIN 384 729  
 CC CHAIN 730 1006  
 CC CHAIN 1007 1615  
 CC CHAIN 1616 1862  
 CC CHAIN 1863 2013  
 CC CHAIN 2014 3010  
 CC TRANSMEM 347 369  
 CC ACT\_SITE 1083 1083  
 CC ACT\_SITE 1107 1107  
 CC ACT\_SITE 1165 1165  
 CC NP\_BIND 1230 1237  
 CC SITE 1316 1319  
 CC CARBOHYD 196 196  
 CC CARBOHYD 209 209  
 CC CARBOHYD 234 234  
 CC CARBOHYD 250 250  
 CC CARBOHYD 305 305  
 CC CARBOHYD 417 417  
 CC CARBOHYD 423 423  
 CC CARBOHYD 430 430  
 CC CARBOHYD 448 448  
 CC CARBOHYD 532 532  
 CC CARBOHYD 556 556

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;  
 Query Match 82.9%; Score 1706; DB 1; Length 3010;  
 Best Local Similarity 78.7%; Pred. No. 5.2e-137;  
 Matches 285; Conservative 39; Mismatches 38; Indels 0; Gaps 0;  
 QY 2 THVTGSGAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYH 614  
 DB 385 THVTGGRVASTQSLVSWLSOGPSQKQLQVNTNGSWHLNRTALNCNDSLQTGFAALEFYA 444  
 QY 62 HKFNSSGCPERLASCRPLTDFDQGWGPISTYANGSGPDQRPYCHYPPKPCGIVPAKSVCG 121  
 DB 445 HRFNASGCPERMASCRPTIDEFAQGWGPIITHDMPSSDQRPYCHYAPRCPGIVPASQVCG 504  
 QY 122 PVYCFPTSPVVGTTDRSGAPTYSGWENDTDVFLVNNTRPPLGNWFGCTWNNSTGFTKVC 181  
 DB 505 PVYCFPTSPVVGTTDRFGAPTYSGWENETDVLNTRPPQGNWFGCTWNNSTGFTKTC 564  
 QY 182 GAPPVIGGAGNNTLHCPITDCFRKHDPATYSRGSGPWITPRCLVDVYPRYLWHYPCITNY 241  
 DB 565 GGPECNIGGVGNNTLVCPITDCFRKHPEATYTKCGSPWLTPRCMVDPYPRYLWHYPCITVNF 624  
 QY 242 TIFKIRMYGVGVEHRLAACAACNWTGRCERCDLDRDRSELSPLLLTTQWQVLPSCFTTLPA 301  
 DB 625 TVFKVMYGVGVEHRLAACAACNWTGRCERCDLDRDRSELSPLLLTTQWQVLPSCFTTLPA 684  
 QY 302 LSTGLIHLHQNIVDVOYLYGVGSSIASWAIKWEYVLLIFLLADARVCSLWMLLLISOA 361  
 DB 685 LSTGLIHLHNRNIVDVOYLYGIGSAVSAFIAKWEYLLIFLLADARVCSLWMLLLISOA 744  
 QY 362 EA 363  
 DB 745 EA 746  
 RESULT 6  
 POLG\_HCVJT  
 ID POLG\_HCVJT STANDARD; PRT; 3010 AA.  
 AC Q00269;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)  
 DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-JT) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92295714; PubMed=1318627;  
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,  
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;  
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese  
 RT carrier: sequence variation within the same individual and among  
 RT infected individuals.";  
 RL Virus Res. 23:39-53(1992).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC  
CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----

DR EMBL: D11168; BAA01943.1; -  
DR PIR: A45573; A45573.  
DR HSP: P27958; 1HEI.  
DR MEROPS: S29.001; -  
DR MEROPS: U39.001; -  
DR INTERPRO: IPR000745; -  
DR INTERPRO: IPR001490; -  
DR INTERPRO: IPR002166; -  
DR INTERPRO: IPR002518; -  
DR INTERPRO: IPR002519; -  
DR INTERPRO: IPR002521; -  
DR INTERPRO: IPR002522; -  
DR INTERPRO: IPR002531; -  
DR INTERPRO: IPR002868; -  
DR PFM: PFM01560; HCV\_NS1; 1.  
DR PFM: PFM01538; HCV\_NS2; 1.  
DR PFM: PFM01006; HCV\_NS4a; 1.  
DR PFM: PFM01001; HCV\_NS4b; 1.  
DR PFM: PFM01506; HCV\_NS5a; 1.  
DR PFM: PFM00998; HCV\_RARP; 1.  
DR PFM: PFM01543; HCV\_capsid; 1.  
DR PFM: PFM01542; HCV\_core; 1.  
DR PFM: PFM01539; HCV\_env; 1.  
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
FT INIT\_MET 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3010  
FT CHAIN 3011 369  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT ACT\_SITE 1230 1230  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 234 234  
FT CARBOHYD 250 250  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;  
  
Query Match 81.2%; Score 1671; DB 1; Length 3010;  
Best Local Similarity 78.5%; Pred. No. 4.9e-134;  
Matches 284; Conservative 29; Mismatches 49; Indels 0; Gaps 0;  
  
QY 2 THVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLSTALNCNDLSNTGLAGLFYH 61  
Db 385 TTTTGGSQARHTQSVTFEFTQGPQRIQLINTNGSHLSTALNCNESLNTGFFAALFYA 444  
QY 62 HKFNSSGCPERLASCRPLTDFOGWGPISYANGSGDPQRPYCHYPPKCGIYPAKSVCG 121  
Db 445 HKFNSSGCPERLASCRPLTDFOGWGPISYANGSGDPQRPYCHYPPKCGIYPAKSVCG 504  
QY 122 PVYCFTPSPVYVGTTRDSGAPYSWGENDTDFVLNTRPPGLNWFCTWMNSTGFTKVC 181  
Db 505 PVYCFTPSPVYVGTTRDSGAPYSWGENDTDFVLNTRPPGLNWFCTWMNSTGFTKVC 564  
QY 182 GAPPVIGGAGNNTLCPTDCERKHPDATYSRCSGSPWITPRCLVDYPRYRLHYPCVTIN 241  
Db 565 GAPPVIGGAGNNTLCPTDCERKHPDATYSRCSGSPWITPRCLVDYPRYRLHYPCVTIN 624  
QY 242 TIFKIRMYGGVEHRLAECNMTNRGERCDLEDRDRSELSPLLLTTTOMQVLPSCFTTLP 301  
Db 625 TIFKIRMYGGVEHRLAECNMTNRGERCDLEDRDRSELSPLLLTTTOMQVLPSCFTTLP 684  
QY 302 LSTGLHLHQNIVDQVLYGVSSIASWAKWYVYVLLLELLADARVCCLNMLLSOA 361  
Db 685 LSTGLHLHQNIVDQVLYGVSSIASWAKWYVYVLLLELLADARVCCLNMLLSOA 744  
QY 362 EA 363  
Db 745 EA 746  
  
RESULT 7  
POLG\_HCVJ8  
ID POLG\_HCVJ8 STANDARD; PRT: 3033 AA.  
AC P26661.  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)  
DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate HC-J8) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92230232; PubMed-1314459;  
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,  
RA Fukuda S., Tsuda F., Mishiro S.;  
RT "Full-length sequence of a hepatitis C virus genome having poor  
RT homology to reported isolates: comparative study of four distinct  
RT genotypes";  
RL Virology 188:331-341(1992).  
CC 1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC 1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL: D00944; BAA00792.1; -
CC PIR: JQ1303; JQ1303.
CC HSP: P27958; 1HEI.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC INTERPRO: IPR000745; -
CC INTERPRO: IPR001490; -
CC INTERPRO: IPR002166; -
CC INTERPRO: IPR002518; -
CC INTERPRO: IPR002519; -
CC INTERPRO: IPR002521; -
CC INTERPRO: IPR002522; -
CC INTERPRO: IPR002531; -
CC INTERPRO: IPR002868; -
CC PFAM: PF01560; HCV_NS1; 1.
CC PFAM: PF01538; HCV_NS2; 1.
CC PFAM: PF01006; HCV_NS4a; 1.
CC PFAM: PF01001; HCV_NS4b; 1.
CC PFAM: PF01506; HCV_NS5a; 1.
CC PFAM: PF00998; HCV_RDRP; 1.
CC PFAM: PF01543; HCV_capsid; 1.
CC PFAM: PF01542; HCV_core; 1.
CC PFAM: PF01539; HCV_env; 1.
CC PFAM: PF01539; HCV_env; 1.
CC KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC KW Core protein; Coat protein; Helicase; ATP-binding;
CC KW Transmembrane; Nonstructural
CC INIT_MET 1 1
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 733
CC CHAIN 734 1010
CC CHAIN 1011 1619
CC CHAIN 1620 1866
CC CHAIN 1867 2017
CC CHAIN 2018 3033
CC CHAIN 3033 369
CC TRANSMEM 347 369
CC ACT_SITE 1087 1087
CC ACT_SITE 1111 1111
CC ACT_SITE 1169 1169
CC NP_BIND 1234 1241
CC SITE 1320 1323
CC CARBOHYD 196 196
CC CARBOHYD 209 209
CC CARBOHYD 234 234
CC CARBOHYD 305 305
CC CARBOHYD 417 417
CC CARBOHYD 423 423
CC CARBOHYD 430 430
CC CARBOHYD 448 448
CC CARBOHYD 477 477
CC CARBOHYD 534 534
CC CARBOHYD 542 542
CC CARBOHYD 558 558
CC CARBOHYD 578 578
CC CARBOHYD 627 627
CC CARBOHYD 649 649
CC CARBOHYD 1091 1091
CC CARBOHYD 2038 2038
CC CARBOHYD 2811 2811
CC SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
```

Query Match

Best Local Similarity 74.4%; Score 1531; DB 1; Length 3033;

Matches 258; Conservative 51; Mismatches 54; Indels 4; Gaps 3;

```

QY 1 ETHVGTGSAGHTVSGFVSLAPGAKQNVQLINTNGSHLSTALNCNDSLNTGWLACLFY 60
DB 384 QTHVGTGSAGHTVSGFVSLAPGAKQNVQLINTNGSHLSTALNCNDSLNTGWLACLFY 443
QY 61 HHKFNSSGCCPERLASCRPLTDFDQGWCPISYA-NGSGP-DQRPYCHYHPKPGCVIPAKS 118
DB 444 THSFNSGCCPERLASCRPLTDFDQGWCPISYA-NGSGP-DQRPYCHYHPKPGCVIPAKS 503
QY 119 VCGPVYCFPTSPVVVGTTRDAGPTYSWGENDTDFVFLNNTPRPLGNWFCCTWNNSTGET 178
DB 504 VCGPVYCFPTSPVVVGTTRDAGPTYSWGENDTDFVFLNNTPRPLGNWFCCTWNNSTGET 563
QY 179 KVCAGPCVITGGAGNNT--LHCPTDFCRKHPDATYSRGSGPWITPRCLVDYYPRLWHYP 236
DB 564 KTCGAPPCRIRADFNASMDLLCPTDCFRKHPDTTYIKCGSGPWLTPRCLLDYYPRLWHYP 623
QY 237 CTINYTIKIRMYVGVGVEHRLAECANWTGRCDELDREDSLSPLLLTTTOHQVLPSCF 296
DB 624 CTINYTIKIRMYVGVGVEHRLAECANWTGRCDELDREDSLSPLLLTTTOHQVLPSCF 683
QY 297 TTLPALSTGLIHLHQNIVDQYLYGVGSSIASWAIKWEYVLLFLLLADARVCSLMMML 356
DB 684 SDLPALSTGLIHLHQNIVDQYLYGVGSSIASWAIKWEYVLLFLLLADARVCSLMMML 743
QY 357 LISQAEA 363
DB 744 LLGQAEA 750

RESULT 9
POLG_HCVJ7 STANDARD; PRT; 737 AA.
AC P27961;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1 AND NS2] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J7) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed-1314459;
RA Okamoto H., Kura K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tanaka F., Mishiro S.;
RT *Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.;
RL Virology 188:331-341(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737
SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 73.3%; Score 1508; DB 1; Length 737;
Best Local Similarity 73.4%; Pred. No. 7.4e-121;
Matches 259; Conservative 37; Mismatches 53; Indels 4; Gaps 2;

QY 2 THVTGSGAGHTVGVSVLLAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLPYH 61
DB 385 TQVTGGAHTVGVSVLLAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLPYH 444
QY 62 HFNSSGCPERLASRPLTDFDQNGPISYANG--SGDPQRYCHYPPKGIYPAKSV 119
DB 445 RFNSSGCPERLSSCKLDDFRIGTLEYENTVNTNEEDMRVCYHYPKPGIYSAKT 504
QY 120 CGPVYCFPTSPVVGTTDRSGAPTSYSGENDTDVFLNTRPPLGNWFGCTWMNSTGFTK 179
DB 505 CGPVYCFPTSPVVGTTDRQVPTYSWGENETDVLNSTRPRAWFGCTWMNSTGFTK 564
QY 180 VCGAPPCVITGGAGNNTLH--CPTDCFRKHPDATYSCSGGPWITPRCLVDYPRYLWHYPC 237
DB 565 TCGAPPCVITGGAGNNTLH--CPTDCFRKHPDATYSCSGGPWITPRCLVDYPRYLWHYPC 624
QY 238 TNYIFIKRMVYGVGVEHLEAACNWTGRCERCDLEDORSELSPLLLTTQWQLPCSF 297
DB 625 TNYIFIKRMVYGVGVEHLEAACNWTGRCERCDLEDORSELSPLLLTTQWQLPCSF 684
QY 298 TLPALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCS 350
DB 685 DLPALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCS 737

RESULT 10
POLG_HCVJ5
ID POLG_HCVJ5 STANDARD; PRT: 737 AA.
AC 27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEINS (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1 AND NS2] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J5) (HCV).

Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.
[1]
SEQUENCE FROM N.A.
MEDLINE=92230232; PubMed=1314459;
Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,
Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
"Full-length sequence of a hepatitis C virus genome having poor
homology to reported isolates: comparative study of four distinct
genotypes.";
RT Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D10075; BAA00969.1;
INTERPRO; IPR002519;
INTERPRO; IPR002521;
INTERPRO; IPR002522;
INTERPRO; IPR002531;
PFAM; PF01560; HCV_NS1; 1.
PFAM; PF01543; HCV_capsid; 1.
PFAM; PF01542; HCV_core; 1.
PFAM; PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
KW INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 >737
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT NON_TER 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CRC64;

Query Match 71.6%; Score 1473; DB 1; Length 737;
Best Local Similarity 71.4%; Pred. No. 7.1e-118;
Matches 252; Conservative 45; Mismatches 52; Indels 4; Gaps 3;

QY 2 THVTGSGAGHTVGVSVLLAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLPYH 61
DB 385 TQVTGGAHTVGVSVLLAPGAKONVOLINTNGSHWLNSTALNCNDSLNTGWLGLPYH 444
```



```
Qy 62 HKFNSSGCPERLASRPLTDFDQGWGPISEA-NGSGP-DQRPYCHWHYPPKCGIYPAKSV 119
Db 445 NRENSGCPHRLSVCRSEAFRIWGLTQYEDNVTNPEDMRPYCHWHYPPKCGIYPAKSV 504
Qy 120 CGPVYCTPSPVVGTTDRSGAPTYSGMENDTVFLNNTNRPPLCGNWFCTWMNSTGFTK 179
Db 505 CGPVYCTPSPVVGTTDARGVPTTYWGENETDVELLNSTRPRGSGNFGCTWMNSTGFTK 564
Qy 180 VCGAPPVIGGAGN-NWTLCHPTDCFRKHPDATYSRGSGPWITPRCLVDYPYRLWHYPC 237
Db 565 TCGAPPCKIRADFNASTDLLCTDCCRKHSDATYIKGSGFWLTPKCMVDYPYRLWHYPC 624
Qy 238 TINYTIKIRMYGVGVEHRLAECNWTGRGERCDLEDRDRSPLLLTWTQVQLPCSF 297
Db 625 TVNYSIFKIRMYGVGVEHRLTAACNFRGDCPNLEDRDRSPLLLTWTQVQLPCSF 684
Qy 298 TLPALSTGLHLHONIVDVQYLVGVGSSIASWATKWEYVVLFLLLADARVCS 350
Db 685 DLPALSTGLHLHONIVDVQYLVGVGSSIASWATKWEYVVLFLLLADARVCA 737

RESULT 11
POLG_HCVH4
ID POLG_HCVH4 STANDARD; PRT: 520 AA.
AC Q01404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93019030; PubMed=1383400;
RA Abe K., Inchauste G., Fujisawa K.;
RT "Genomic characterization and mutation rate of hepatitis C virus
RT isolated from a patient who contracted hepatitis during an epidemic
RT of non-A, non-B hepatitis in Japan.";
RL J. Gen. Virol. 73:2725-2729(1992).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10688; BAA01530.1;
CC DR INTERPRO: IPR002519;
CC DR INTERPRO: IPR002521;
CC DR INTERPRO: IPR002522;
CC DR INTERPRO: IPR002531;
CC DR PFAM: PF01560; HCV_Ns1; 1;
CC DR PFAM: PF01543; HCV_capsid; 1;
CC DR PFAM: PF01542; HCV_core; 1;
CC DR PFAM: PF01539; HCV_env; 1;
CC KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC INIT_MET 1
CC REMOVED FROM CAPSID PROTEIN C BY THE
CC CELLULAR AMINOPEPTIDASE.
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 >520
CC CHAIN 520 347
CC TRANSMEM BY SIMILARITY.
```

```
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 520 AA; 56499 MW; AAL35246CF20D525 CRC64;

Query Match 27.0%; Score 555; DB 1; Length 520;
Best Local Similarity 68.9%; Pred. No. 6.2e-40;
Matches 93; Conservative 12; Mismatches 30; Indels 0; Gaps 0;

Qy 2 THVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHILNSTALNCNDSLNTGLAGLFYH 61
Db 386 THVTGSGAGHTVSGFVSLAPGAKQNVOLINTNGSWHILNSTALNCNDSLNTGLAGLFYH 61
Qy 62 HKFNSSGCPERLASRPLTDFDQGWGPISEANGSGPDRPYCHWHYPPKCGIYPAKSVCG 121
Db 446 HKFNSSGCPERLASRPLTDFDQGWGPISEANGSGPDRPYCHWHYPPKCGIYPAKSVCG 121
Qy 122 PVYCTPSPVVGTT 136
Db 506 PVYCTPSPVVGTT 520

RESULT 12
POLG_HCVJ2
ID POLG_HCVJ2 STANDARD; PRT: 513 AA.
AC P27959;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate HC-J2) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kuri K., Okada S.I., Yamamoto K., Lizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes";
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10074; BAA00968.1;
CC DR INTERPRO: IPR002519;
CC DR INTERPRO: IPR002521;
CC DR INTERPRO: IPR002522;
CC DR INTERPRO: IPR002531;
CC DR PFAM: PF01560; HCV_Ns1; 1.
```

DR PFAM: PF01543; HCV\_capsid; 1.  
 DR PFAM: PF01542; HCV\_core; 1.  
 DR PFAM: PF01539; HCV\_env; 1.  
 DR KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
 DR Transmembrane; Nonstructural protein.  
 DR FT INIT\_MET 1 1  
 DR CHAIN 1 115  
 DR CHAIN 116 191  
 DR CHAIN 192 383  
 DR CHAIN 384 >513  
 DR TRANSMEM 347 369  
 DR CARBOHYD 196 196  
 DR CARBOHYD 209 209  
 DR CARBOHYD 233 233  
 DR CARBOHYD 234 234  
 DR CARBOHYD 250 250  
 DR CARBOHYD 305 305  
 DR CARBOHYD 417 417  
 DR CARBOHYD 423 423  
 DR CARBOHYD 430 430  
 DR CARBOHYD 448 448  
 DR NON\_TER 513  
 DR SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 26.9%; Score 553; DB 1; Length 513;  
 Best Local Similarity 72.1%; Pred. No. 9e-40; Mismatches 10; Indels 0; Gaps 0;  
 Matches 93; Conservative 10; Mismatches 26; Indels 0; Gaps 0;  
 QY 2 THVTGSGAGTGVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSINTGWLGLFYH 61  
 DB 385 THVTGATGHTTSGTASLFLPGASOKIQLINTNGSHLNTALNCNDSINTGFLAALFY 444  
 QY 62 HKNSSGCPERLASCRPLTDFDQGWGPIYSYANGSGDORPCYHPPKPCGIVPAKSVCG 121  
 DB 445 HKFNASGCPERLASCRSIDFGDQGWGPIYTFEPGSDQKPYCHYAPQRCVSYSAADVCG 504  
 QY 122 PVYCETPSP 130  
 DB 505 PVYCETPSP 513

RESULT 13  
 POLG\_HCVHK  
 ID POLG\_HCVHK STANDARD; PRT; 520 AA.  
 AC Q01403;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1)] (FRAGMENT).  
 OS Hepatitis C virus (isolate HCV-KF) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93019030; PubMed=1383400;  
 RA Abe K., Inchauspe G., Fujisawa K.;  
 RT "Genomic characterization and mutation rate of hepatitis C virus  
 RT isolated from a patient who contracted hepatitis during an epidemic  
 RT of non-A, non-B hepatitis in Japan."  
 RL J. Gen. Virol. 73:2725-2729(1992).  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND MRNA.  
 CC PROTEIN C AND MRNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: D10687; BAA01529.1; --  
 DR PIR: JQ1925; JQ1925.  
 DR INTERPRO: IPR002519; --  
 DR INTERPRO: IPR002521; --  
 DR INTERPRO: IPR002522; --  
 DR INTERPRO: IPR002531; --  
 DR PFAM: PF01560; HCV\_NSI; 1.  
 DR PFAM: PF01543; HCV\_capsid; 1.  
 DR PFAM: PF01542; HCV\_core; 1.  
 DR PFAM: PF01539; HCV\_env; 1.  
 DR KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
 DR Transmembrane; Nonstructural protein.  
 DR FT INIT\_MET 1 1  
 DR CHAIN 1 115  
 DR CHAIN 116 191  
 DR CHAIN 192 383  
 DR CHAIN 384 >520  
 DR TRANSMEM 347 369  
 DR CARBOHYD 196 196  
 DR CARBOHYD 209 209  
 DR CARBOHYD 233 233  
 DR CARBOHYD 234 234  
 DR CARBOHYD 305 305  
 DR CARBOHYD 418 418  
 DR CARBOHYD 424 424  
 DR CARBOHYD 431 431  
 DR CARBOHYD 449 449  
 DR NON\_TER 520  
 DR SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 26.6%; Score 548; DB 1; Length 520;  
 Best Local Similarity 68.1%; Pred. No. 2.4e-31;  
 Matches 92; Conservative 12; Mismatches 31; Indels 0; Gaps 0;  
 QY 2 THVTGSGAGTGVSGFVSLAPGAKONVOLINTNGSHLNTALNCNDSINTGWLGLFYH 61  
 DB 386 THVTGTEGFATRLTSLFALGPSKIQIINTNGSHLNTALNCNDSFKTGLAALFY 445  
 QY 62 HKNSSGCPERLASCRPLTDFDQGWGPIYSYANGSGDORPCYHPPKPCGIVPAKSVCG 121  
 DB 446 HKFNASGCPERHMASCRPIDKFDQGWGPIYAPPSISEQRCYCHYAPRCPTIPASEVCG 505  
 QY 122 PVYCETPSPVVVGTT 136  
 DB 506 PVYCETPSPVVVGTT 520

RESULT 14  
 POLG\_HCVH8  
 ID POLG\_HCVH8 STANDARD; PRT; 321 AA.  
 AC Q27956;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: MATRIX PROTEIN (ENVELOPE PROTEIN M);  
 DE MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN NS1] (FRAGMENT).  
 OS Hepatitis C virus (isolate HCV18) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91112009; PubMed=1846505;  
 RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,  
 RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,  
 RA Han J.H.;  
 RT "Variable and hypervariable domains are found in the regions of HCV  
 RT corresponding to the flavivirus envelope and NS1 proteins and the

pestivirus envelope glycoproteins.";  
Virology 180:842-848(1991).  
-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
PROTEIN C AND MRNA.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X53131; CAA37291.1; -  
DR INTERPRO: IPR002519; -  
DR INTERPRO: IPR002531; -  
DR PFAM: PF01560; HCV\_NSI; 1.  
DR PFAM: PF01542; HCV\_core; 1.  
DR PFAM: PF01539; HCV\_env; 1.  
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
KW Transmembrane; Nonstructural protein.  
FT NON\_TER 1  
FT CHAIN 1  
FT CHAIN 77 267  
FT CHAIN 268 >321  
FT CARBOHYD 80 80  
FT CARBOHYD 93 93  
FT CARBOHYD 118 118  
FT CARBOHYD 189 189  
FT CARBOHYD 301 301  
FT CARBOHYD 307 307  
FT CARBOHYD 314 314  
FT CARBOHYD 321 321  
FT NON\_TER 321  
SQ SEQUENCE 321 AA; 34238 MW; 2F5DE79F7C7845C8 CRC64;  
  
Query Match 11.6%; Score 238; DB 1; Length 321;  
Best Local Similarity 75.9%; Pred. No. 3.4e-13;  
Matches 41; Conservative 6; Mismatches 7; Indels 0; Gaps 0;  
  
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGW 54  
II: :|||:||||:| ||| ||||| |||||:|||||:|||||:|||||  
Db 268 ETVTSGNAGHTMTGIVRFAPGPKQNVHLINTNGSWHLNSTALNCNDSINTGW 321  
  
RESULT 15  
POLG\_HCVTH STANDARD; PRT; 321 AA.  
AC P27957;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE GENOME POLYPEPTIDE [CONTAINS: MATRIX PROTEIN (ENVELOPE PROTEIN M);  
DE MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN NSI] (FRAGMENT).  
OS Hepatitis C virus (Isolate TH) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91112009; PubMed-1846505;  
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,  
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,  
RA Han J.H.;  
RT "Variable and hypervariable domains are found in the regions of HCV  
corresponding to the flavivirus envelope and NS1 proteins and the  
RT pestivirus envelope glycoproteins.";  
RL Virology 180:842-848(1991).  
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

## PROTEIN C AND MRNA.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X53131; CAA37291.1; -  
DR INTERPRO: IPR002519; -  
DR INTERPRO: IPR002531; -  
DR PFAM: PF01560; HCV\_NSI; 1.  
DR PFAM: PF01542; HCV\_core; 1.  
DR PFAM: PF01539; HCV\_env; 1.  
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
KW Transmembrane; Nonstructural protein.  
FT NON\_TER 1  
FT CHAIN 1  
FT CHAIN 76 267  
FT CHAIN 268 >321  
FT CARBOHYD 80 80  
FT CARBOHYD 93 93  
FT CARBOHYD 118 118  
FT CARBOHYD 189 189  
FT CARBOHYD 301 301  
FT CARBOHYD 307 307  
FT CARBOHYD 314 314  
FT CARBOHYD 321 321  
FT NON\_TER 321  
SQ SEQUENCE 321 AA; 34074 MW; B2EB83F521C3B520 CRC64;

Query Match 10.9%; Score 225; DB 1; Length 321;  
Best Local Similarity 75.9%; Pred. No. 4.3e-12;  
Matches 41; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGW 54  
II: :|||||:| ||| ||||| |||||:|||||:|||||:|||||  
Db 268 ETVTGGSAHAGALGIASLFNQGARQIQLINTNGSWHLNSTALNCNDSINTGW 321

Search completed: March 6, 2001, 11:53:43  
JOB time: 160 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:54:53 ; Search time 75.33 Seconds  
(without alignments)  
564.801 Million cell updates/sec

Title: US-09-407-430-2  
Perfect score: 2058  
Sequence: 1 ETHVTGSGAGTSGFVSL.....ADARVCSLWMLLSQAEA 363

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_15:\*

- 1: sp-archaea:\*
- 2: sp-bacteria:\*
- 3: sp-fungi:\*
- 4: sp-human:\*
- 5: sp-invertebrate:\*
- 6: sp-mammal:\*
- 7: sp-mhc:\*
- 8: sp-organelle:\*
- 9: sp-phage:\*
- 10: sp-plant:\*
- 11: sp-rodent:\*
- 12: sp-virus:\*
- 13: sp-vertebrate:\*
- 14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2046	99.4	3011	12 Q91FE5	Q91FE5 hepatitis c
2	1957	95.1	778	12 Q04184	Q04184 h genome po
3	1949	94.7	3011	12 Q36579	Q36579 hepatitis c
4	1946	94.6	3011	12 Q36608	Q36608 hepatitis c
5	1943	94.4	3011	12 Q03463	Q03463 hepatitis c
6	1942	94.4	3011	12 Q36610	Q36610 hepatitis c
7	1932	93.9	3011	12 Q36609	Q36609 hepatitis c
8	1925	93.5	778	12 Q04185	Q04185 h genome po
9	1817	88.3	403	12 Q9PX22	Q9PX22 hepatitis c
10	1794	87.2	3010	12 Q9J3H0	Q9J3H0 hepatitis c
11	1790	87.0	3011	12 Q81754	Q81754 hepatitis c
12	1745	84.8	3010	12 P90194	P90194 hepatitis c
13	1744	84.7	3010	12 Q81760	Q81760 hepatitis c
14	1744	84.7	3010	12 Q9J3H9	Q9J3H9 hepatitis c
15	1742	84.6	3010	12 Q02828	Q02828 h genome po
16	1742	84.6	3010	12 Q9J3G1	Q9J3G1 hepatitis c
17	1741	84.6	2864	12 Q92975	Q92975 hepatitis c
18	1741	84.6	3010	12 Q9J3G8	Q9J3G8 hepatitis c
19	1740	84.5	2864	12 Q9WLK8	Q9WLK8 hepatitis c

20	1740	84.5	3010	12	Q92969	Q92969 hepatitis c
21	1740	84.5	3010	12	Q81989	Q81989 hepatitis c
22	1740	84.5	3013	12	Q9J3H4	Q9J3H4 hepatitis c
23	1739	84.5	3010	12	Q02829	Q02829 h genome po
24	1738	84.5	3010	12	Q9Q1V6	Q9Q1V6 hepatitis c
25	1737	84.4	2864	12	Q9WLK0	Q9WLK0 hepatitis c
26	1737	84.4	3010	12	Q92970	Q92970 hepatitis c
27	1737	84.4	3010	12	Q93016	Q93016 hepatitis c
28	1736	84.4	3010	12	Q9J3G4	Q9J3G4 hepatitis c
29	1735	84.3	3010	12	Q9J310	Q9J310 hepatitis c
30	1733	84.2	2864	12	Q92976	Q92976 hepatitis c
31	1732	84.2	3015	12	Q9WBH5	Q9WBH5 hepatitis c
32	1731	84.1	3010	12	Q9Q1X4	Q9Q1X4 hepatitis c
33	1731	84.1	3010	12	Q9Q1X2	Q9Q1X2 hepatitis c
34	1731	84.1	3010	12	Q9J3H8	Q9J3H8 hepatitis c
35	1730	84.1	3010	12	Q9Q1X5	Q9Q1X5 hepatitis c
36	1730	84.1	3010	12	Q9J3G2	Q9J3G2 hepatitis c
37	1729	84.0	3010	12	Q9J3G9	Q9J3G9 hepatitis c
38	1728	84.0	802	12	Q68520	Q68520 hepatitis c
39	1728	84.0	2864	12	Q9WLK9	Q9WLK9 hepatitis c
40	1728	84.0	3010	12	Q92971	Q92971 hepatitis c
41	1728	84.0	3010	12	P90192	P90192 hepatitis c
42	1728	84.0	3010	12	Q9Q6P1	Q9Q6P1 hepatitis c
43	1728	84.0	3010	12	Q9J3H1	Q9J3H1 hepatitis c
44	1727	83.9	2864	12	Q92974	Q92974 hepatitis c
45	1727	83.9	3010	12	Q9J3H3	Q9J3H3 hepatitis c

## ALIGNMENTS

## RESULT 1

Q91FE5  
ID Q91FE5 PRELIMINARY; PRT; 3011 AA.  
AC Q91FE5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.;  
RT "An Infectious Clone of the HCV-1 Prototype Sequence";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF271632; AAF81759.1;  
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;

Query Match 99.4%; Score 2046; DB 12; Length 3011;  
Best Local Similarity 99.4%; Pred. No. 9,8e-189;  
Matches 361; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	ETHVTGSGAGTSGFVSLAPGAKQNVOLINTNGSHLNTALNCNDSLTGWLGLFY	60
Db	384	ETHVTGSGAGTSGFVSLAPGAKQNVOLINTNGSHLNTALNCNDSLTGWLGLFY	443
QY	61	HKFNSSGCPERLASCRLTDFDQGWGPISYANGSGDPDPCYWHYPPKCGIYPAKSVC	120
Db	444	HKFNSSGCPERLASCRLTDFDQGWGPISYANGSGDPDPCYWHYPPKCGIYPAKSVC	503
QY	121	GPVYCFPTSPVVVGTDRSGAPTYSWGENDTDFVLNNTRPPLGNWFCTWMNSTGTFTKV	180
Db	504	GPVYCFPTSPVVVGTDRSGAPTYSWGENDTDFVLNNTRPPLGNWFCTWMNSTGTFTKV	563
QY	181	CGAPPCVIGGAGNNTLHCPTDCFRKHDPDATYSRCSGSPWTPRCLVDYPRVRLWHYPCIN	240
Db	564	CGAPPCVIGGAGNNTLHCPTDCFRKHDPDATYSRCSGSPWTPRCLVDYPRVRLWHYPCIN	623
QY	241	YTIKIRMYGVGVEHLEACNWTGRGERCDLEDRSLSPLLLTTTQWQVLPSCFTTLP	300

```

Db 624 YTFKIRMYVGGVEHRELEAACNWTGRGCDLEDRSELSPLLLTTQWVLPSCFTTLP 683
QY 301 ALSTGLIHLHQNIVDYQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 360
Db 684 ALSTGLIHLHQNIVDYQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 2
Q04184 PRELIMINARY; PRT; 778 AA.
AC Q04184; Q81810;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C
(CORE PROTEIN); MATRIX PROTEIN
(ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN
NS1; NONSTRUCTURAL PROTEIN NS2; NONSTRUCTURAL PROTEIN NS4A;
NONSTRUCTURAL PROTEIN NS4B; HELICASE
(NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)] (FRAGMENT).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91195357; PubMed=1849654;
RA Oyata N., Alter H.J., Miller R.H., Purcell R.H.;
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C
virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3
AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND MRNA.
CC EMBL; M62381; AAB02127.1; -.
DR INTERPRO; IPR002518; -.
DR INTERPRO; IPR002519; -.
DR INTERPRO; IPR002521; -.
DR INTERPRO; IPR002522; -.
DR INTERPRO; IPR002523; -.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
KW Polyprotein; Glycoprotein; RNA-directed RNA polymerase; Core protein;
Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane;
Nonstructural protein.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
FT CHAIN 2 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2
(POTENTIAL).
FT CHAIN 730 >778 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT NON_TER 778 778
SQ SEQUENCE 778 AA; 85143 MW; F7709172CD03E39B CRC64;

Query Match 95.1%; Score 1957; DB 12; Length 778;
Best Local Similarity 94.2%; Pred. No. 8.3e-181;
Matches 342; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

```

```

QY 1 ETHVTGSSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSINTGWLGLFY 60
Db 384 ETHVTGSSAGRTTAGLVGLLTPGAKQNIQINTNGSWHLNSTALNCNDSINTGWLGLFY 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGGPISYANGSPDORPCYCHWYPPKPCGIVPAKSVC 120
Db 444 HHKFNSSGCPERLASCRRLTDFDQGGPISYANGSGLDERPCYCHWYPPKPCGIVPAKSVC 503
QY 121 GPVYCFTSPVGVVGTDRSGAPTSYWGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180
Db 504 GPVYCFTSPVGVVGTDRSGAPTSYWGANDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 563
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYDYPYRLWHYPCIN 240
Db 564 CGAPPCVIGGAGNNTLLCPTDCFRKHPDATYSRCGSGPWITPRCMVDYDYPYRLWHYPCIN 623
QY 241 YTFKIRMYVGGVEHRELEAACNWTGRGCDLEDRSELSPLLLTTQWVLPSCFTTLP 300
Db 624 YTFKIRMYVGGVEHRELEAACNWTGRGCDLEDRSELSPLLLTTQWVLPSCFTTLP 683
QY 301 ALSTGLIHLHQNIVDYQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 360
Db 684 ALSTGLIHLHQNIVDYQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 743
QY 361 AEA 363
Db 744 AEA 746

RESULT 3
Q036579 PRELIMINARY; PRT; 3011 AA.
AC Q036579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
transcribed RNA.";
RL Science 277:570-574(1997).
DR EMBL; AF009606; AAB66324.1; -.
DR INTERPRO; IPR000745; -.
DR INTERPRO; IPR001410; -.
DR INTERPRO; IPR001490; -.
DR INTERPRO; IPR002166; -.
DR INTERPRO; IPR002518; -.
DR INTERPRO; IPR002519; -.
DR INTERPRO; IPR002521; -.
DR INTERPRO; IPR002522; -.
DR INTERPRO; IPR002531; -.
DR INTERPRO; IPR002868; -.
DR PFAM; PF00998; HCV_RdRP; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01506; HCV_NS5a; 1.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
DR PRODOM; PD186062; -. 1.
KW Polyprotein.

```

RESULT	4	
OC36608		
ID	OC36608	PRELIMINARY; PRT: 3011 AA.
AC	OC36608;	
DT	01-JAN-1998	(TREMBLrel. 05, Created)
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)
DE	POLYPROTEIN.	
OS	Hepatitis C virus strain H77.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus.	
OX	NCBI_TaxID=63746;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RP	STRAIN=H77;	
RC	MEDLINE=97385173; PubMed=9238047;	
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;	
RT	"Transcripts from a single full-length cDNA clone of hepatitis C virus	
RT	are infectious when directly transfected into the liver of a	
RT	RT chimpanzee.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).	
DR	EMBL; AF011751; AAB67036.1; -;	
DR	INTERPRO; IPR000745; -;	
DR	INTERPRO; IPR001410; -;	
DR	INTERPRO; IPR001490; -;	
DR	INTERPRO; IPR002166; -;	
DR	INTERPRO; IPR002518; -;	
DR	INTERPRO; IPR002519; -;	
DR	INTERPRO; IPR002521; -;	
DR	INTERPRO; IPR002522; -;	
DR	INTERPRO; IPR002531; -;	
DR	INTERPRO; IPR002868; -;	
DR	PFAM; PF00998; HCV_RdRp; 1.	
DR	PFAM; PF01001; HCV_NS4b; 1.	
DR	PFAM; PF01006; HCV_NS4a; 1.	
DR	PFAM; PF01506; HCV_NS5a; 1.	
DR	PFAM; PF01538; HCV_NS2; 1.	

RESULT	5
Q03463	
ID	Q03463 PRELIMINARY; PRT: 3011 AA.
AC	Q03463;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE	GENOME POLYPROTEIN.
OS	Hepatitis C virus.
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX	Hepacivirus.
NCBI_TaxID=11103;	
[1]	SEQUENCE FROM N.A.
RX	MEDLINE=93117120; PubMed=1335573;
RA	Okamoto H., Kanai N., Mishiro S.;
RT	"Full-length nucleotide sequence of a Japanese hepatitis C virus isolate (HC-J1) with high homology to USA isolates.";
RL	Nucleic Acids Res. 20:6410-6410(1992).
[2]	
RN	SEQUENCE FROM N.A.
RX	MEDLINE=91013116; PubMed=2170712;
RA	Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA	Yoshizawa H., Tsuda F., Miyakawa Y., Mayumi M.;
RT	"The 5'-terminal sequence of the hepatitis C virus genome.";
RL	Jpn. J. Exp. Med. 60:167-177(1990).
DR	EMBL: D10749; BAA01582.1; -.
DR	HSSP: P27958; IHEI.
DR	INTERPRO: IPR000745; -.
DR	INTERPRO: IPR001410; -.
DR	INTERPRO: IPR001490; -.





[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=H77;  
RX MEDLINE=97385173; PubMed=9238047;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus  
RT are infectious when directly transfected into the liver of a  
RT chimpanzee.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).  
DR EMBL: AF011752; AAB67037.1; -;  
DR HSSP: P27938; IHEI.  
DR INTERPRO: IPR000745; -;  
DR INTERPRO: IPR001410; -;  
DR INTERPRO: IPR001490; -;  
DR INTERPRO: IPR002166; -;  
DR INTERPRO: IPR002518; -;  
DR INTERPRO: IPR002519; -;  
DR INTERPRO: IPR002521; -;  
DR INTERPRO: IPR002522; -;  
DR INTERPRO: IPR002531; -;  
DR INTERPRO: IPR002868; -;  
DR PFAM: PF00998; HCV\_RdRP; 1.  
DR PFAM: PF01001; HCV\_NS4b; 1.  
DR PFAM: PF01006; HCV\_NS4a; 1.  
DR PFAM: PF01506; HCV\_NS4a; 1.  
DR PFAM: PF01538; HCV\_NS4a; 1.  
DR PFAM: PF01539; HCV\_NS2; 1.  
DR PFAM: PF01542; HCV\_core; 1.  
DR PFAM: PF01543; HCV\_core; 1.  
DR PFAM: PF01560; HCV\_capsid; 1.  
DR PRODOM: PD186062; -; 1.  
KW Polyprotein.  
SQ SEQUENCE 3011 AA; 327262 MW; 10D1C9702CA9B5DC CRC64;

Query Match 93.9%; Score 1932; DB 12; Length 3011;  
Best Local Similarity 93.4%; Pred. No. 1e-177;  
Matches 339; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLNTGLAGLFY 60  
Db 384 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLNTGLAGLFY 443

Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPCYCHYPPKPGCIVPAKSV 120  
Db 444 RHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDLDRPCYCHYPPKPGCIVPAKSV 503

Qy 121 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 504 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563

Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 240  
Db 564 RVPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 623

Qy 241 YTFKIRMYVGGVEHLEAACNWTGERCDLDRDRSELSPLLLTTQVQLPCSFITLP 300  
Db 624 YTFKIRMYVGGVEHLEAACNWTGERCDLDRDRSELSPLLLTTQVQLPCSFITLP 683

Qy 301 ALSTGLIHLHQNIVDQVLYGVGSSIASWAIKWEYVLLFLLLADARVCSCLLMMLLISQ 360  
Db 684 ALSTGLIHLHQNIVDQVLYGVGSSIASWAIKWEYVLLFLLLADARVCSCLLMMLLISQ 743

Qy 361 AEA 363  
Db 744 AEA 746

RESULT 8  
Q04185  
ID Q04185 PRELIMINARY; PRT: 778 AA.  
AC Q04185: Q01811;  
DT 01-NOV-1998 (Tremblrel. 08, Created)

DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C  
DE (CORE PROTEIN); MATRIX PROTEIN  
DE (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN  
DE NS1; NONSTRUCTURAL PROTEIN NS2; NONSTRUCTURAL PROTEIN NS4B;  
DE NONSTRUCTURAL PROTEIN NS4B; HELICASE  
DE (NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)] (FRAGMENT).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91195357; PubMed=1849654;  
RA Ogata N., Alter H.J., Miller R.H.;  
RT "Nucleotide sequence and mutation rate of the H strain of hepatitis C  
RT virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).  
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3  
CC AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
DR EMBL: M62382; AAB02128.1; -;  
DR INTERPRO: IPR002518; -;  
DR INTERPRO: IPR002519; -;  
DR INTERPRO: IPR002521; -;  
DR INTERPRO: IPR002522; -;  
DR INTERPRO: IPR002531; -;  
DR PFAM: PF01538; HCV\_NS2; 1.  
DR PFAM: PF01539; HCV\_env; 1.  
DR PFAM: PF01542; HCV\_core; 1.  
DR PFAM: PF01543; HCV\_capsid; 1.  
DR PFAM: PF01560; HCV\_NS1; 1.  
KW Polyprotein; Glycoprotein; RNA-directed RNA polymerase; Core protein;  
KW Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane;  
KW Nonstructural protein.  
FT INIT\_MET 1  
FT CHAIN 2 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 >778  
FT TRANSMEM 347 369  
FT NON\_TER 778 778  
FT SEQUENCE 778 AA; 85530 MW; 6E37E50AB97B6C32 CRC64;

Query Match 93.5%; Score 1925; DB 12; Length 778;  
Best Local Similarity 91.7%; Pred. No. 1e-177;  
Matches 333; Conservative 13; Mismatches 17; Indels 0; Gaps 0;

Qy 1 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLNTGLAGLFY 60  
Db 384 ETHVTVGSAGHTVSGFVSLAPGAKQNVOLINTNGSWHLNLTALNCNDSLNTGLAGLFY 443

Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPCYCHYPPKPGCIVPAKSV 120  
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPCYCHYPPKPGCIVPAKSV 503

Qy 121 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 504 GPVYCFPSVVGTTDRSGAPTSYANGNDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563

Qy 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 240  
Db 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 623

```
QY 241 YTIKIRMYGGVEHRLAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 624 YTIKIRMYGGVEHRLAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWMLLSQ 360
DB 684 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746

RESULT 9
ID Q9PXZ2 PRELIMINARY; PRT; 403 AA.
AC Q9PXZ2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE E2/NSI REGION.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92348860; Pubmed=1668329;
RA Kresdorf D., Porchon C., Brechot C.;
RT "Hepatitis C virus (HCV)-RNA in non-A, non-B chronic hepatitis in
J. France. Nucleotide sequence of a French HCV isolate.";
RL J. Hepatol. 13:24-24(1991).
DR INTERPRO; IPR002519;
DR INTERPRO; IPR002531;
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01560; HCV_NSI; 1.
SQ SEQUENCE 403 AA; 44565 MW; 06278192EAC5B3F1 CRC64;
```

```
Query Match 88.3%; Score 1817; DB 12; Length 403;
Best Local Similarity 86.7%; Pred. No. 1.3e-167;
Matches 306; Conservative 30; Mismatches 17; Indels 0; Gaps 0;
```

```
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFY 60
DB 51 EYITGGSTARTQGLVSLFSGAKQDILQINTNGSWHLNSTALNCNDSLTGWLGLFY 110
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHPKPGIVPAKSVC 120
DB 111 YHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHPKPGIVPAQTV 170
QY 121 GPVYCFTSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIVPAKSVC 180
DB 171 GPVYCFTSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIVPAQTV 230
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWITPRCLVDYPIRLWHYPCIN 240
DB 231 CGAPPCVIGGAGNNTLYCPTDCFRKHDPATYSRCGSGPWITPRCLVDYPIRLWHYPCIN 290
QY 241 YTIKIRMYGGVEHRLAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 291 YTIKIRMYGGVEHRLAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 350
QY 301 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWMLLSQ 353
DB 351 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWMLLSQ 403
```

```
RESULT 10
ID Q9J3H0 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
```

```
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ND22;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF207763; AAF65953.1;
SQ SEQUENCE 3010 AA; 327098 MW; 737EEF31E3C2B28D CRC64;
```

```
Query Match 87.2%; Score 1794; DB 12; Length 3010;
Best Local Similarity 83.5%; Pred. No. 2.3e-164;
Matches 303; Conservative 30; Mismatches 30; Indels 0; Gaps 0;
```

```
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFY 60
DB 384 DTHVTGGVYAHQTQGFASLFPAGASQNIQLINTNGSWHLNSTALNCNDSLTGWLGLFY 443
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCHWHPKPGIVPAKSVC 120
DB 444 THEFNASSGCPERMASCRPIDQFQGWGPIYTEPNSDPQRPYCHWHPKPGIVPAKSVC 503
QY 121 GPVYCFTSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIVPAKSVC 180
DB 504 GPVYCFTSPVVVGTDRSGAPTSYANGSGDPQRPYCHWHPKPGIVPAKSVC 563
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWITPRCLVDYPIRLWHYPCIN 240
DB 564 CGAPPCVIGGAGNNTLHCPTDCFRKHDPATYSRCGSGPWITPRCLVDYPIRLWHYPCIN 623
QY 241 YTIKIRMYGGVEHRLAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300
DB 624 YTIKIRMYGGVEHRLAACNWTGRGCDLEDRDRSELSPLLLTTTQWQVLPSCFTTLP 683
QY 301 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWMLLSQ 360
DB 684 ALSTGLHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMWMLLSQ 743
QY 361 AEA 363
DB 744 AEA 746
```

```
RESULT 11
ID Q81754 PRELIMINARY; PRT; 3011 AA.
AC Q81754;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RA Okamoto H.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RX MEDLINE=94172337; Pubmed=8126459;
RA Okamoto H., Kojima M., Sakamoto M., Iizuka H., Hadiwandowo S.,
```



	Query Match	84.7%	Score 1744	DB 12	Length 3010
	Best Local Similarity	80.9%	Pred. No. 1.5e-159		
	Matches 293	Conservative 35	Mismatches 34	Indels. 0	Gaps 0
QY	2	THVTGGSAGHTVVSGFVLLAPCAKQNVQLINTNGSWHLNSTALNCDSINTGWLAGLFVH	61		
		: : : :	: : : :	: : : :	: : : :
DB	385	TYVTGGAAARGASGTTSLFSRGSQKILVNTNGSWHLNRTALNCDSNFTGFLAALFYA	444		
		: : : :	: : : :	: : : :	: : : :
	62	HKFNSGGCPERLASCRPLTDFDQGWGPTSYANGSGPDQRPYCHWYPPRCGTIVPAKSYCG	121		
QY		: : : :	: : : :	: : : :	: : : :

Query Match	84.78;	Score 1744;	DB 12;	Length 3010;
Best Local Similarity	81.58;	Pred. No. 1.5e-159;		
Matches 296;	Conservative 32;	Mismatches 35;	Indels 0;	Gaps 0;
QY	1	ETHVTGSAGHTVSGFVLLAPGAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLAFY	60	
DB	384	ETRVTGGAARATSGFASLERSGSSQIOLINTNGSWHINTALNCNDSLOTGFALAFY	443	
QY	61	HHKNSGCCPRLRLASCRLPTDFDQGWGPISTANGSGPQRPYCWHPKPCGIVPAKSWC	120	
DB	444	THKTNASGCPERMASCRSIDKFDQGWGITTAEPNSDQPYCWHYAPRQCQIVTASQVC	503	
QY	121	GPVYCFTPSPVVVGTTRDSGAPTVSWGENDTDFVLYNTRPPLGNWFGCTWNNSTGFTKV	180	
DB	504	GPVYCFTPSPVVVGTTRDFGVTYTVSWGNETDVLUNTRPPQGNWFGCTWNNSTGFTKT	563	
QY	181	CGAPPCVTGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPIRLWHYPCTTN	240	
DB	564	CGPPCDIGGVGNNTLTCTDCFRKHPEATYTKCGSGPWLTPRCLVNYPIRYLWHYPCTVN	623	
QY	241	YTIKIRMYVGGVEHRLAACNWTGERCDLDRDRSELSPLLLTTTQWVLPSCFTTLP	300	
DB	624	FTIPKVRMYVGGVEHRLAACNWTGERCDLDRDRSELSPLLLSTTEWVILPCSFSTTLP	683	
QY	301	ALSTGLIHLHQNIVDQVLYGVGSSIAWAIKWEEVYVLLFLLADARVCSCLWMMLLIQ	360	
DB	684	ALSTGLIHLHNRNIVDQVLYIGSVASWFAIKWEEVYVLLFLLADARVCACLWMMLLIAQ	743	





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:51:45 ; Search time 47.68 Seconds  
(without alignments)  
182.157 Million cell updates/sec

Title: US-09-407-430-3

Perfect score: 1494

Sequence: 1 GAKNVQLINTGSHLNST.....LEACNWTGRCDELDLDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1494	100.0	402	14 R34438	Sequence of glycoprotein
2	1494	100.0	480	14 R33992	HCV-1 E2/NSI prote
3	1494	100.0	663	17 R92935	HCV1 E2 + NS2 poly
4	1494	100.0	663	20 W67615	Hepatitis C virus
5	1494	100.0	2772	11 R08123	Hepatitis C virus
6	1494	100.0	2955	20 Y14975	Amino acid sequenc
7	1494	100.0	3011	13 R21519	Compiled HCV sequ
8	1494	100.0	3011	17 R90931	Hepatitis C virus
9	1494	100.0	3011	18 W34480	HCV polyprotein.
10	1494	100.0	3011	19 W40038	HCV polyprotein.
11	1492	99.9	3011	14 R31621	Hepatitis C virus
12	1489	99.7	2955	11 R08124	Hepatitis C virus

13	1486	99.5	2816	14 R34009	HCV-1 polyprotein.
14	1473	98.6	621	14 R33185	Sequence of subfra
15	1473	98.6	622	14 R33591	HCV CKS-NS1S1-NS1S
16	1473	98.6	738	14 R33592	HCV CKS-full lengt
17	1473	98.6	2894	13 R24440	Composite HCV HC-2
18	1473	98.6	2894	16 R70230	Composite hepatitis
19	1438	96.3	3011	15 R66995	Hepatitis C virus
20	1426	95.4	402	14 R34439	Sequence of glycop
21	1426	95.4	409	14 R33995	H77 E2/NSI protein
22	1426	95.4	733	14 R38278	NANB hepatitis vir
23	1419	95.0	1006	18 W12715	HCV genome type 1a
24	1418	94.9	3011	19 W77397	Hepatitis C virus
25	1418	94.9	3011	20 W98020	Infectious hepatitis
26	1416	94.8	3011	19 W77398	Hepatitis C virus
27	1414	94.6	402	14 R34440	Sequence of glycop
28	1414	94.6	409	14 R33996	H90 E2/NSI protein
29	1413	94.6	305	18 W00929	Recombinant HCV E2
30	1413	94.6	333	14 R40118	HGH-HCV-E2 fusion
31	1413	94.6	367	16 R79218	pHCV167-encoded pr
32	1413	94.6	359	14 R40117	HGH-HCV-E2 fusion
33	1413	94.6	537	14 R40117	APP-HCV-E2 fusion
34	1413	94.6	967	16 R79222	pHCV141-encoded se
35	1413	94.6	1648	16 R79221	pHCV176-encoded se
36	1413	94.6	3011	14 R40119	HCV genomic amino
37	1413	94.6	3011	16 R79232	HCV sequence. Hep
38	1407	94.2	367	14 R40115	APP-HCV-E2 fusion
39	1397	93.5	3011	14 R40120	HCV genomic amino
40	1390	93.0	3011	13 R22154	NANBV Hutch c59 is
41	1389	93.0	278	14 R33997	Th E2/NSI protein.
42	1386	92.8	337	16 R79217	pHCV351-encoded AP
43	1386	92.8	377	16 R79226	pHCV422-encoded AP
44	1386	92.8	397	16 R79220	pHCV419-encoded AP
45	1386	92.8	410	16 R79227	pHCV423-encoded AP

#### ALIGNMENTS

RESULT 1	
R34438	
ID R34438 standard; Protein; 402 AA.	
XX	
AC R34438;	
XX	
DT 09-AUG-1993 (first entry)	
XX	
DE Sequence of glycoprotein E2/NSI in clone HCV1.	
XX	
KW Hepatitis C virus; envelope protein; glycoprotein; E2/NSI;	
KW diagnostic reagent.	
XX	
OS Hepatitis C virus.	
XX	
PN EP537626-A.	
XX	
PD 21-APR-1993.	
XX	
PF 08-OCT-1992; 92EP-0117191.	
XX	
PR 08-OCT-1991; 91JP-0260824.	
XX	
PA (NAHE-) NAT INST OF HEALTH.	
XX	
PI Harada S, Honda Y, Miyamura T, Saito I;	
XX	
DR WPI; 1993-127516/16.	
DR N-PSDB; Q40330.	
XX	
PT Diagnostic reagent for hepatitis C virus - comprises second	
PT envelope protein or first non-structural protein encoded by HCV	
XX	
PS gene and has sugar chain	
PS Claim 2; Pages 30-32; 58pp; English.	

```

XX CC Glycoprotein E2/NS1 is derived from the second envelope protein or
CC first non-structural protein encoded by the genome of HCV. The
CC nucleic acid is extracted from the serum of the patient of hepatitis
CC C. The serum is pref. mixed with transfer RNA (tRNA) as a carrier
CC of virus RNA. As a technique of cloning cDNA from the nucleic acid,
CC it is preferred to use polymerase chain reaction method. In the
CC reaction, any commercially available random primers or synthesized
CC DNA having a base sequence similar to that of primer AS1 may be used
CC as a primer. Representative examples of sense primers include S1.
XX CC
XX SQ Sequence 402 AA;

Query Match 100.0%; Score 1494; DB 14; Length 402;
Best Local Similarity 100.0%; Pred. No. 7.8e-125; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYYHHKFNSSGCPERLASCRPLTDF 60
DB 67 gakqvqlntngswhlntstalnncndslntgwlglfyyhhkfnssgcperscrpltdf 126

QY 61 DQGWGPISYANGSGDQRPYCYWHYPKPGCIVPAKSVCGPVYCFTPSPVVVGTTRSGAP 120
DB 127 dqgwgpiyangsgdqrpycwyhpyppkpgcivpaksvcgpvycftpspvvvgtttdrsgap 186

QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWNNSTGFTKVCAPPVCVIGAGNNTLHCPTDC 180
DB 187 tyswgendtDFVNLNTRPPLGNWFGCTWNNSTGFTKVCAPPVCVIGAGNNTLHCPTDC 246

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRLWHYPCTINYTIKIRMYVGGVEHRLAECN 240
DB 247 frkhpdatysrcsgsgpwitprclvdyprylwhypctinytifkirmvvgvghrleaecn 306

QY 241 WTRGERCDLEDRDR 254
DB 307 wtrgercdleldr 320

RESULT 2
R33992 ID R33992 standard; Protein; 480 AA.
XX AC
XX AC R33992;
XX DT 26-JUL-1993 (first entry)
XX DE HCV-1 E2/NS1 protein.
XX KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;
XX KW HCV; asymptomatic; chronically infected; epitope; viral isolate;
XX KW domain; immunological; cross-reactive; envelope protein; vaccine;
XX KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.
XX OS Synthetic.
XX OS WO9306126-A.
XX PN 01-APR-1993.
XX PD 11-SEP-1992; 92WO-US07683.
XX PF 13-SEP-1991; 91US-0759575.
XX PR (CHIR ) CHIRON CORP.
XX PA Houghton M, Weiner AJ;
XX PI WPI; 1993-117468/14.
XX DR Immuno-reactive hepatitis C virus polypeptide compsns. - contg.
XX PT at least 2 sequences from the first variable domain of distinct
XX PT HCV isolates

```

```

XX PS Disclosure; Fig 3; 106pp; English.
XX CC
XX CC The sequences given in R33992-002 represent a portion of the E2/NS1
CC protein encoded by group I and group II HCV isolates, from amino acid
CC 370-850. E2/NS1 are viral envelope proteins and are of immunogenic
CC interest. E2/NS1 contains an N-terminal hypervariable domain of about
CC 30 amino acids which shows large variation between nearly all isolates.
CC This is an important immunoreactive domain. This putative envelope
CC glycoprotein E2/NS1 may correspond to the gp53(BVDV)/gp55 (hog cholera
CC virus) envelope polypeptide of the pestiviruses and the NS1 of the
CC flaviviruses, both of which confer protective immunity in hosts
CC vaccinated with these polypeptides. It has been discovered that a
CC number of important HCV epitopes vary among viral isolates and that
CC these epitopes can be mapped to specific domains. This meant that
CC immunologically cross-reactive polypeptides which focus on variable
CC rather than constant domains can be produced. See also Q39134-48
CC and R33982-91.
XX CC
XX SQ Sequence 480 AA;

Query Match 100.0%; Score 1494; DB 14; Length 480;
Best Local Similarity 100.0%; Pred. No. 9.6e-125; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;

QY 1 GAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYYHHKFNSSGCPERLASCRPLTDF 60
DB 37 gakqvqlntngswhlntstalnncndslntgwlglfyyhhkfnssgcperscrpltdf 96

QY 61 DQGWGPISYANGSGDQRPYCYWHYPKPGCIVPAKSVCGPVYCFTPSPVVVGTTRSGAP 120
DB 97 dqgwgpiyangsgdqrpycwyhpyppkpgcivpaksvcgpvycftpspvvvgtttdrsgap 156

QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWNNSTGFTKVCAPPVCVIGAGNNTLHCPTDC 180
DB 157 tyswgendtDFVNLNTRPPLGNWFGCTWNNSTGFTKVCAPPVCVIGAGNNTLHCPTDC 216

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRLWHYPCTINYTIKIRMYVGGVEHRLAECN 240
DB 217 frkhpdatysrcsgsgpwitprclvdyprylwhypctinytifkirmvvgvghrleaecn 276

QY 241 WTRGERCDLEDRDR 254
DB 277 wtrgercdleldr 290

RESULT 3
R33992 ID R33992 standard; Protein; 663 AA.
XX AC
XX AC R33992;
XX DT 04-JUN-1996 (first entry)
XX DE HCV1 E2 + NS2 polypeptide.
XX KW HCV1; E1 protein; E2 protein; truncation; vaccine; diagnosis;
XX KW therapy; protein secretion.
XX OS Hepatitis c virus.
XX FH Key Location/Qualifiers
XX FT Peptide 1..21 /label= Sig_peptide
XX FT Protein 22..336 /label= Mat_protein
XX FT Domain 337..361 /note= "E2 mature protein"
XX FT /label= Membrane-spanning_domain
XX FT /note= "E2 C-terminal anchor"
XX FT Protein 362..663 /label= Mat_protein

```



/note= "NS2A+NS2B proteins"

FT XX W09604301-A2.  
 PN XX 15-FEB-1996.  
 XX XX 26-JUL-1995; 95WO-US10035.  
 XX XX 25-JUL-1995; 95US-0282959.  
 PR XX 29-JUL-1994; 94US-0282959.  
 XX XX (CHIR ) CHIRON CORP.  
 PA XX Houghton M, Selby M;  
 XX WPI: 1996-129331/13.  
 DR N-PSDB; T18246.  
 XX Secretatable hepatitis C virus E1 and E2 polypeptide(s) lacking  
 PT all/part of the membrane spanning domain - useful in vaccines, and  
 PT for diagnostic and therapeutic purposes, e.g. in assays for HCV  
 XX PS Disclosure: Fig 4a-4c; 46pp; English.  
 CC Truncated versions of the hepatitis C virus E2 envelope protein  
 CC (R92935) are obtd. by deletion the membrane-spanning C-terminal  
 CC anchor domain (approx. amino acids 337-361) of the protein.  
 CC This can be achieved by PCR amplification of the E2 gene (T18246)  
 CC using appropriate primers. Truncated versions of the E1  
 CC polypeptide (see R92934) are similarly produced. When produced  
 CC recombinantly in host cells, the truncated proteins are secreted  
 CC into the medium. When co-expressed or combined after separate  
 CC expression, the truncated E2 and E1 proteins form a complex. The  
 CC truncated proteins and complex are useful in vaccines and for  
 CC diagnostic and therapeutic appls.  
 XX SQ Sequence 663 AA;

Query Match 100.0%; Score 1494; DB 17; Length 663;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-124;  
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 DB 43 gaknvqlintngshlnstalnclndslntglaglfyhhkfnssgcperscrpltdf 102  
 QY 61 DQGWGPISYANGSGDPQRPYCHYHPKPGCIIPAKSVCGPVYCFPTSPVVVGTDRSGAP 120  
 DB 103 dggwgpisyangsgdpqrpychwhppkpcgipaksvcgpvycftspvvvgttdrsgap 162  
 QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPDTC 180  
 DB 163 tyswgendtdvflntrpplgnwfgctwmnstgftkvcgappcviggagnntlhcpdtdc 222  
 QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCPTINVTIFKIRMYVGVGVEHLEAACN 240  
 DB 223 frkhdpatysrcsgspwltprclvdypyrwlhwhypctinytlfkirmvvgvghrleaacn 282  
 QY 241 WTRGERCDLEDRDR 254  
 DB 283 wtrgercdledrdr 296

RESULT 4  
 W67615  
 ID W67615 standard; Protein; 663 AA.  
 XX AC W67615;  
 XX DT 02-MAR-1999 (first entry)  
 XX DE Hepatitis C virus E2 protein.

XX Hepatitis C virus; E2; HCV; truncation; variant; PCR; amplification;  
 KW affinity chromatography; Galanthus nivalis; agglutinin; medicament;  
 KW detection; infection.  
 XX Hepatitis C virus.  
 OS W09850556-A2.  
 PN 12-NOV-1998.  
 XX 06-MAY-1998; 98WO-US09097.  
 PF 06-MAY-1997; 97US-0045675.  
 PR (CHIR ) CHIRON CORP.  
 XX PA Abridgani S, Chien D, Choo QL, Glazer E, Houghton M;  
 PI Selby M;  
 XX WPI: 1999-034724/03.  
 DR N-PSDB; V81371.  
 XX Methods for isolating truncated HCV E1 and E2 polypeptides - used  
 PT in, e.g. immunodiagnostic kits for diagnosis of HCV infection  
 XX PS Disclosure: Fig 2A-E; 65pp; English.  
 CC This sequence represents the Hepatitis C virus (HCV) E2 protein. The  
 CC invention relates to the intracellular production and isolation of  
 CC C-terminally truncated variants of the E2 protein. The truncations  
 CC start from about amino acid residue 500 of the E2 protein and are  
 CC generated by PCR amplification of the gene sequence. The truncated  
 CC proteins are then produced intracellularly in host cells which are  
 CC disrupted. The HCV proteins are then purified preferably by affinity  
 CC chromatography, especially using Galanthus nivalis agglutinin resin.  
 CC The HCV E1 and HCV E2 polypeptides can be used to manufacture a  
 CC medicament useful for detecting the presence or absence of HCV infection  
 CC in an individual. They can also be used in a immunodiagnostic test kit  
 CC for detecting HCV infection.  
 XX SQ Sequence 663 AA;

Query Match 100.0%; Score 1494; DB 20; Length 663;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-124;  
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 DB 43 gaknvqlintngshlnstalnclndslntglaglfyhhkfnssgcperscrpltdf 102  
 QY 61 DQGWGPISYANGSGDPQRPYCHYHPKPGCIIPAKSVCGPVYCFPTSPVVVGTDRSGAP 120  
 DB 103 dggwgpisyangsgdpqrpychwhppkpcgipaksvcgpvycftspvvvgttdrsgap 162  
 QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPDTC 180  
 DB 163 tyswgendtdvflntrpplgnwfgctwmnstgftkvcgappcviggagnntlhcpdtdc 222  
 QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCPTINVTIFKIRMYVGVGVEHLEAACN 240  
 DB 223 frkhdpatysrcsgspwltprclvdypyrwlhwhypctinytlfkirmvvgvghrleaacn 282  
 QY 241 WTRGERCDLEDRDR 254  
 DB 283 wtrgercdledrdr 296

RESULT 5  
 R08123  
 ID R08123 standard; protein; 2772 AA.  
 XX

AC R08123;  
XX 23-JAN-1991 (first entry)  
XX Hepatitis C virus polypeptide from long ORF.  
DE Hepatitis C virus; antiviral agent.  
XX Hepatitis C virus.  
XX Hepatitis C virus.  
XX EP388232-A.  
XX 19-SEP-1990.  
XX 16-MAR-1990; 90EP-0302866.  
XX 18-MAY-1989; 89US-0355002.  
XX 17-MAR-1989; 89US-0325338.  
XX 20-APR-1989; 89US-0341334.  
XX (CHIR-) CHIRON CORP.  
XX Houghton M, Choo QL, Kuo G;  
XX WPI; 1990-284418/38.  
XX N-PSDB; Q05955.  
XX Hepatitis C virus DNA - used for producing probes,  
XX polypeptide(s), antibodies and anti-sense polynucleotide(s) for  
XX diagnosis and therapy.  
XX Disclosure; Fig 16; 83pp; English.  
XX HCV cDNA libraries were constructed using pooled serum from a  
XX chimpanzee with chronic HCV infection. A lambda gill library was  
XX screened with probes derived from previously isolated clones. The  
XX ORF is derived from the overlapping clones p14a, CA167b, CA156e,  
XX CA84a, CA59a, K9-1, 12f, 14i, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36,  
XX 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g and 15e.  
XX This polypeptide can be used to design probes  
XX for the detection of HCV nucleic acids, in screening programmes  
XX for antiviral agents and in preparing blood free of HCV. Antisense  
XX polynucleotides can be used to inhibit viral replication.  
XX See also Q05956.  
XX Sequence 2772 AA;  
SQ  
Query Match 100.0%; Score 1494; DB 11; Length 2772;  
Best Local Similarity 100.0%; Pred. No. 7.6e-124;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAKNQVLINTNGSWHLNSTALNCNDSINTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60  
Db 292 gaknqvqlintngswhlnstalnncndsintgwlglaglfyhhkfnssgpcperlasrpltdf 351  
Qy 61 DOGWGPISYANGSGDQRPYCHWYPPKPGCIVPAKSVCGPVYCTPSPVVGTTDRSGAP 120  
Db 352 dggwgpisyangsgdqrpychwyppkpgcivpaksvcgpyctpspvvgttdrsgap 411  
Qy 121 TYSWGENDTDFVNLNTRPPLGNFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180  
Db 412 tyswgendtdvfnlntprplgnfgctwmnstgtfkcvcgappcvigagantlhcptdc 471  
Qy 181 FRKHPDARYSCGSPWITPRLCLVDYDYPRLWHYDCTINTYTFKIRMYGVGVEHLEACN 240  
Db 472 frkhpdayrscgspwitprclvdydypylwhypctintytfkirmvgygvehleacn 531  
Qy 241 WTRGERCDLEDRDR 254  
Db 532 wtrgercdleddr 545

RESULT 6  
Y14975  
ID Y14975 standard; Protein; 2955 AA.  
XX  
AC Y14975;  
XX 08-NOV-1999 (first entry)  
XX Amino acid sequence of HCV-1 ORF.  
XX Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;  
XX HCV infection; vaccine.  
XX Hepatitis C virus.  
XX Key Location/Qualifiers  
XX Misc-difference 441 /note= "encoded by TT"  
XX Misc-difference 461 /note= "encoded by CCCC"  
XX EP939128-A2.  
XX 01-SEP-1999.  
XX 17-SEP-1990; 90EP-0310149.  
XX 21-DEC-1989; 89US-0456142.  
XX 15-SEP-1989; 89US-0408045.  
XX (CHIR ) CHIRON CORP.  
XX (OYAA/) OYA A.  
XX Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;  
XX Miyamura T, Saito I, Weiner AJ;  
XX WPI; 1999-480843/41.  
XX N-PSDB; 207656.  
XX New Hepatitis C Virus isolates, useful for diagnosis of hepatitis  
XX infections and development of vaccines  
XX Disclosure; Fig 12; 132pp; English.  
XX The invention provides two new isolates of hepatitis C virus (HCV), J1  
XX and J7. These two isolates comprise nucleotide and amino acid sequences  
XX that are distinct from the HCV isolate HCV-1. The nucleotide sequences  
XX may be used to detect non-A, non-B HCV (NANBH) polynucleotides by  
XX hybridisation for diagnosis of NANBH infections. They may also be used to  
XX screen blood donors, donated blood and blood products for this infection.  
XX The isolates may also be used to isolate other naturally occurring  
XX variants of the virus. The polypeptides may be used as a vaccine for  
XX administration to patients to protect against infection with NANBH. The  
XX present sequence represents the amino acid sequence of HCV-1 ORF.  
XX Sequence 2955 AA;  
SQ  
Query Match 100.0%; Score 1494; DB 20; Length 2955;  
Best Local Similarity 100.0%; Pred. No. 8.2e-124;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAKNQVLINTNGSWHLNSTALNCNDSINTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60  
Db 406 gaknqvqlintngswhlnstalnncndsintgwlglaglfyhhkfnssgpcperlasrpltdf 465  
Qy 61 DOGWGPISYANGSGDQRPYCHWYPPKPGCIVPAKSVCGPVYCTPSPVVGTTDRSGAP 120  
Db 466 dggwgpisyangsgdqrpychwyppkpgcivpaksvcgpyctpspvvgttdrsgap 525  
Qy 121 TYSWGENDTDFVNLNTRPPLGNFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180  
Db 526 tyswgendtdvfnlntprplgnfgctwmnstgtfkcvcgappcvigagantlhcptdc 585

QY 181 FRKHPDATYSCGSGPWITPRCLVDYPRRLWHYPCTINTYIFKIRMYGVGVEHLEAACN 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 586 frkhpdatysrcsgsgpwitprclvdyprrlwhypctintyifkirmvvgvhehrleaacn 645  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 241 WTRGERCDLEDRDR 254  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 646 wtrgercdledrdr 659

RESULT 7  
 R21519  
 ID R21519 standard; Protein; 3011 AA.

XX AC R21519;

XX DT 22-JUN-1992 (first entry)

XX XX Compiled HCV sequence.

XX KW HCV1; serum; gt11.

XX OS Hepatitis C virus 1.

XX FH Key Location/Qualifiers

FT MISC-difference 9 /label= ARG  
 FT MISC-difference 11 /label= THR  
 FT MISC-difference 176 /label= THR  
 FT MISC-difference 334 /label= VAL  
 FT MISC-difference 603 /label= ILE  
 FT MISC-difference 848 /label= (ASN)  
 FT MISC-difference 1114 /label= SER  
 FT MISC-difference 1117 /label= THR  
 FT MISC-difference 1276 /label= LEU  
 FT MISC-difference 1328 /label= (VAL)  
 FT MISC-difference 1454 /label= TYR  
 FT MISC-difference 1471 /label= (SER)  
 FT MISC-difference 1877 /label= (GLY)  
 FT MISC-difference 1948 /label= (HIS)  
 FT MISC-difference 1949 /label= (CYS)  
 FT MISC-difference 2021 /label= (VAL)  
 FT MISC-difference 2349 /label= (SER)  
 FT MISC-difference 2385 /label= (PHE)  
 FT MISC-difference 2386 /label= (ALA)  
 FT MISC-difference 2502 /label= (PHE)  
 FT MISC-difference 2690 /label= (GLY)  
 FT MISC-difference 2996 /label= (PRO)  
 XX WO202642-A.  
 XX 20-FEB-1992.

XX PF 12-AUG-1991; 91WO-US05728.  
 XX PR 10-AUG-1990; 90US-0566209.  
 XX PA (CHIR-) CHIRON CORP.  
 XX PI Houghton M, Choo QL, Kuo G, Weiner AJ, Urdea MS, Irvine BD;  
 PI Kolberg JA;  
 XX DR WPI; 1992-080094/10.  
 XX DR N-PSDB; Q21744.  
 XX PT Reagents for isolating, amplifying and detecting HCV  
 PT polynucleotide(s) - used to monitor spread of blood-borne non-a,  
 PT non-b hepatitis virus infection and screen blood samples for  
 PT virus  
 XX PS Disclosure; Fig 1; 67pp; English.  
 XX CC Heterogeneities in cloned DNAs of HCV1 are indicated by the amino  
 CC acid indicated in the features, the parentheses indicated that the  
 CC heterogeneity was detected at or near to the 5' - or 3'-end of the  
 CC HCV in the clone.  
 CC The sequence is derived from a composite HCV cDNA from HCV1, a  
 CC prototypic HCV. The DNA sequence is based upon sequence information  
 CC derived from a no. of HCV cDNA clones, which were isolated from a no.  
 CC of HCV cDNA libraries, including the "c" library present in lambda  
 CC gt11 (ATCC No.40394), and from human serum. The HCV cDNA clones  
 CC were isolated by methods described in WO9014436.  
 CC The clones from which the sequence was derived are 5'clone32,  
 CC b114a, 18g, ag30a, CA205a, CA250a, CA216a, p14a, CA167b, CA156e,  
 CC CA84a, CA59a, K9-1 (also called K9-1), 26j, 131, 12f, 14i, 11b, 7f,  
 CC 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g,  
 CC 39c, 35f, 19g, 26g, 15e, b5a, 16jh, 6k and p131jh.  
 XX SQ Sequence 3011 AA;

Query Match 100.0%; Score 1494; DB 13; Length 3011;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-124;  
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSCCPERLASCRPLTDF 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 406 gakqnvqlintngswhlNSTALNCNDSLNTGWLGLAGLFYHHKFNSSCCPERLASCRPLTDF 465  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 61 DQGWGPISYANGSGDPDRPYCWHYPPKCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 466 dggwgpisyangsgdpdrpycwhypkpcgylpaksvcgpcvfcftspvvvgtdrsgap 525  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 121 TYSWGENDTDFVLNNTRPLGNWFGCTWMNSTGTGFTKVCAGPCVIGGAGNNTLHCPTDC 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 526 tyswgenndtfdvlnntrpplgnwfgctwmnstgtfckvcgappcvlggagntllhcptdc 585  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 181 FRKHPDATYSCGSGPWITPRCLVDYPRRLWHYPCTINTYIFKIRMYGVGVEHLEAACN 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 586 frkhpdatysrcsgsgpwitprclvdyprrlwhypctintyifkirmvvgvhehrleaacn 645  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 241 WTRGERCDLEDRDR 254  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 646 wtrgercdledrdr 659

RESULT 8  
 R90931  
 ID R90931 standard; Protein; 3011 AA.  
 XX AC R90931;  
 XX DT 15-MAY-1996 (first entry)  
 XX DE Hepatitis C virus polypeptide.

XX Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection;  
 KW diagnosis; antibodies.

XX Hepatitis C virus.

XX Key Location/Qualifiers  
 FT Misc-difference 1..122  
 FT /label= antigen  
 FT /note= "C22; R90936"  
 FT Misc-difference 199..328  
 FT /label= antigen  
 FT /note= "S2; R90935"  
 FT Misc-difference 1192..1457  
 FT /label= antigen  
 FT /note= "C33c; R90932"  
 FT Misc-difference 1569..1931  
 FT /label= antigen  
 FT /note= "C100; R90933"  
 FT Misc-difference 2054..2464  
 FT /label= antigen  
 FT /note= "NS5; R90934"

XX EP693687-Al.

XX 24-JAN-1996.

XX 03-APR-1991; 91EP-0114016.

XX 04-APR-1990; 90US-0504352.

XX (CHIR ) CHIRON CORP.

XX Choo Q, Houghton M, Kuo G;

XX WPI; 1996-117956/13.

XX N-PSDB; T12710.

XX Combinations of synthetic Hepatitis C Virus antigens - provide more  
 PT effective diagnosis of Non-A, Non-B Hepatitis

XX Disclosure; Fig 1(A-Y): 53pp; English.

XX The combination comprises an HCV antigen from the C domain (pref.  
 CC C22 - R90936) and at least one HCV antigen from the NS3 (pref. C33c  
 CC - R90932), NS4 (pref. C100 - R90933), S (pref. S2 - R90935) or NS5  
 CC (R90934) domain.

XX The antigens may in the form of a fusion protein, a simple physical  
 CC mixture, or the individual antigens commonly bound to a solid matrix.  
 CC They are pref. prepd. by recombinant DNA techniques (primers are  
 CC given in T12711-T12716), but can be synthesised or isolated from  
 CC HCV using affinity chromatography.

XX Sequence 3011 AA;

Query Match 100.0%; Score 1494; DB 17; Length 3011;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-124;  
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHLSTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
 DB 406 gakqvqlintngshlntalnncnslntgwlglfyyhkhfnssgcpelrascrlptdf 465  
 QY 61 DCGWPIYANGSGDQRPYCHWHYPPKCGIVPAKSVGCVYCFPTSPVYVGTTRDSGAP 120  
 DB 466 dggwpiyangsgdqrpychwhypkpcgilypaksvcgvyccftpsvvygttdrsgap 525  
 QY 121 TYSWGENDTDFVLNTRPPLGNFTGCTWMNSTGTFTKVCGAPPCVIGAGNNTLHCPTDC 180  
 DB 526 tyswgenndtfdvlntrpplgnftgctwmnstgtftkvcgappcvigagntnlhcptdc 585  
 QY 181 FRKHDPATYSCSGFWITPRCLVDYPRVRLWHYPCNTINTYIKIRMYVGGVEHRLAECN 240

DB 586 frkhpdatysrcsgfwitprclvdypyrilwhypctintytkirmyvggvehrleaecn 645  
 QY 241 WTRGERCDLEDRDR 254  
 DB 646 wtrgercdledrdr 659  
 RESULT 9  
 W34480  
 ID W34480 standard; Protein; 3011 AA.  
 XX AC W34480;  
 XX 16-MAR-1998 (first entry)  
 XX HCV polyprotein.  
 DE PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;  
 KW C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;  
 KW NS4.  
 XX Hepatitis C virus.

XX Key Location/Qualifiers  
 FT Misc-difference 366 /note= "can optionally be Arg"  
 FT Misc-difference 372 /note= "can optionally be Thr"  
 FT Misc-difference 867 /note= "can optionally be Thr"  
 FT Misc-difference 1341 /note= "can optionally be Val"  
 FT Misc-difference 2148 /note= "can optionally be Ile"  
 FT Misc-difference 2883 /note= "can optionally be Asn"  
 FT Misc-difference 3681 /note= "can optionally be Ser"  
 FT Misc-difference 3690 /note= "can optionally be Thr"  
 FT Misc-difference 4167 /note= "can optionally be Leu"  
 FT Misc-difference 4323 /note= "can optionally be Val"  
 FT Misc-difference 4701 /note= "can optionally be Tyr"  
 FT Misc-difference 4732 /note= "can optionally be Ser"  
 FT Misc-difference 5970 /note= "can optionally be Gly"  
 FT Misc-difference 6183 /note= "can optionally be His"  
 FT Misc-difference 6186 /note= "can optionally be Cys"  
 FT Misc-difference 6402 /note= "can optionally be Val"  
 FT Misc-difference 7386 /note= "can optionally be Ser"  
 FT Misc-difference 7494 /note= "can optionally be Phe"  
 FT Misc-difference 7497 /note= "can optionally be Ala"  
 FT Misc-difference 7845 /note= "can optionally be Phe"  
 FT Misc-difference 8409 /note= "can optionally be Gly"  
 FT Misc-difference 9102 /note= "can optionally be Gly"  
 FT Misc-difference 9327 /note= "can optionally be Pro"  
 XX US5683864-A.

```

XX PD 04-NOV-1997.
XX PF 07-JUL-1992; 92US-0910760.
XX PR 07-JUL-1992; 92US-0910760.
PR 18-NOV-1987; 87US-0122714.
PR 30-DEC-1987; 87US-0139886.
PR 26-FEB-1988; 88US-0161072.
PR 06-MAY-1988; 88US-0191263.
PR 26-OCT-1988; 88US-0263584.
PR 14-NOV-1988; 88US-0271450.
PR 17-MAR-1989; 89US-0325338.
PR 20-APR-1989; 89US-0341334.
PR 21-APR-1989; 89US-0353896.
PR 04-APR-1990; 90US-0504332.
XX PA (CHIR ) CHIRON CORP.
XX PI Choo Q, Houghton M, Kuo G;
XX DR WPI: 1997-548976/50.
XX DR N-PSDB; T99981.
XX PT Combination of three hepatitis C virus antigens - used for detection
XX PS of specific antibodies to diagnose infection
XX PS Disclosure; Column 25-46; 57pp; English.
XX CC This sequence represents the Hepatitis C virus polyprotein. Fragments of
CC the DNA encoding this sequence can be amplified and used in the
CC combination of HCV antigens of the invention. The HCV antigen combination
CC comprises an antigen (Ag1) comprising the C domain (i.e. amino acids (aa)
CC 1-120 of the HCV polyprotein), or its immunologically reactive fragment
CC containing at least 8 aa. It also comprises two additional antigens from
CC two different polyprotein domains, including at least 8 aa from the NS3,
CC NS4, S or NS5 domains of the polyprotein, corresponding, respectively, to
CC aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein.
CC Alternatively, Ag1 contains at least 8 aa from the 1-122 or 9-177 aa
CC regions of the HCV polyprotein. These antigen combinations are used
CC diagnostically to detect anti-HCV antibodies, using any standard
CC immunoassay format. These antigen combinations have a broader range of
CC reactivity with antibodies than any antigen individually.
XX SQ Sequence 3011 AA;

Query Match 100.0%; Score 1494; DB 18; Length 3011;
Best Local Similarity 100.0%; Pred. NO. 8.3e-124;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTGSHWLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
Db 406 gAkqnvqlIntgshwlnstalnCNdSLntglAGlfYhHkfnSSGCPERlasCRPltdf 465
QY 61 DQWGPISYANGSGPDQRPYCHYHPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 120
Db 466 dQwgpIsyAngsgpDqRpyChyHPKpCGIvPaKsvCGpVYcFTpSPvVGtTDRsGAP 525
QY 121 TYSWGENDDTVFLNTRPLGNWFECTWNNSTGTFKVCAPPCVIGGAGNNTLHCPTDC 180
Db 526 tYswgEnDdTvFlNtrPlGnWfEcTWNnStGtFkVcApPCvIGgAGnNtLhCpTdc 585
QY 181 FRKHPDATYSRCSGSPWITPRCLVDYFYRLWHYPCTINVTIFKIRMYGVGVEHRLAECN 240
Db 586 frkHpDatysrCSgspWiTprClvDYfYrLwhYpCTINvTIFkIRMyGVgVEhRLeACN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 wtrGerCDleDRdr 659

RESULT 10

```

```

W40038
ID W40038 standard; Protein; 3011 AA.
XX AC W40038;
XX DT 26-MAY-1998 (first entry)
XX DE HCV polyprotein.
XX KW Hepatitis C virus C domain; HCV; C antigen; immunological activity;
KW NS3 domain; NS4 domain; S domain; NS5 domain.
XX OS Hepatitis C virus.
XX FH Key
FH Domain 1..120
FH Modified-site 9
FH /label= C_domain
FH /note= "As given in the specification this amino
FT acid can also be Arg"
FH Modified-site 11
FT /note= "As given in the specification this amino
FT acid can also be Thr"
FH Domain 120..400
FH Modified-site 174
FT /label= S_domain
FT /note= "As given in the specification this amino
FT acid can also be Thr"
FH Modified-site 334
FT /note= "As given in the specification this amino
FT acid can also be Val"
FH Modified-site 603
FT /note= "As given in the specification this amino
FT acid can also be Ile"
FH Modified-site 847
FT /note= "As given in the specification this amino
FT acid can also be Asn"
FH Domain 1050..1640
FH Modified-site 1114
FT /label= NS3_domain
FT /note= "As given in the specification this amino
FT acid can also be Ser"
FH Modified-site 1217
FT /note= "As given in the specification this amino
FT acid can also be Thr"
FH Modified-site 1276
FT /note= "As given in the specification this amino
FT acid can also be Leu"
FH Modified-site 1328
FT /note= "As given in the specification this amino
FT acid can also be Val"
FH Modified-site 1452
FT /note= "As given in the specification this amino
FT acid can also be Tyr"
FH Modified-site 1472
FT /note= "As given in the specification this amino
FT acid can also be Ser"
FH Domain 1640..2000
FH Modified-site 1877
FT /label= NS4_domain
FT /note= "As given in the specification this amino
FT acid can also be Gly"
FH Modified-site 1948
FT /note= "As given in the specification this amino
FT acid can also be His"
FH Modified-site 1949
FT /note= "As given in the specification this amino
FT acid can also be Cys"
FH Domain 2000..3011
FH Modified-site 2021
FT /label= NS5_domain
FT /note= "As given in the specification this amino
FT acid can also be Val"

```

FT	Modified-site	2348	/note=	"As given in the specification this amino acid can also be Ser"
FT				
FT	Modified-site	2385	/note=	"As given in the specification this amino acid can,also be Phe"
FT				
FT	Modified-site	2386	/note=	"As given in the specification this amino acid can also be Ala"
FT				
FT	Modified-site	2502	/note=	"As given in the specification this amino acid can also be Phe"
FT				
FT	Modified-site	2690	/note=	"As given in the specification this amino acid can also be Gly"
FT				
FT	Modified-site	2921	/note=	"As given in the specification this amino acid can also be Gly"
FT				
FT	Modified-site	2996	/note=	"As given in the specification this amino acid can also be Pro"
FT				
XX		US5712087-A.		
PN				
XX				
XX				
PD		27-JAN-1998.		
XX				
PF		12-MAY-1995;	95US-0440519.	
XX				
XX		07-JUL-1992;	92US-0910760.	
PR		04-APR-1990;	90US-0504352.	
PR		12-MAY-1995;	95US-0440519.	
XX				
PA		(CHIR ) CHIRON CORP.		
XX				
PI		Choo Q, Houghton M, Kuo G;		
XX				
XX		WPI: 1998-119973/11.		
DR		N-PSDB; V09989.		
XX				
PT		Immunoassays for hepatitis C virus antibodies - using combinations of antigenic fragments of HCV polyprotein		
PT				
XX				
PS		Disclosure; Fig 1; 59pp; English.		
XX				
CC		This sequence represents the hepatitis C virus (HCV) polyprotein which is used in the construction of novel combinations of HCV antigens that have a broader range of immunological activity than any single HCV antigen. An example of such an antigen given in this specification comprises a first antigen containing at least 8 amino acids of the C domain of the HCV polyprotein and a second antigen comprising at least 8 amino acids of the NS3 domain, the NS4 domain, the S domain or the NS5 domain of the HCV polyprotein in the form of a fusion protein, a physical mixture or bound to a solid matrix.		
CC		Note: The features given in the specification as represented in the feature table of W40038 differ from the positions indicated in Figure 1.		
XX		Sequence	3011 AA;	
SO				

	Query Match	100.0%	Score 1494;	DB 19;	Length 3011;
	Best Local Similarity	100.0%;	Pred. No. 8.3e-124;		
	Matches 254;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GAKQNVQLINTGSHLNSTALNCNDSLNTGWLAGLFYHHKFNSSGCPERLASCRPLTDF	60		
Db	406	gakqnvqlintgshlntalnccndslntgwlaglfyhkhfnssgperlascrpltdf	465		
Qy	61	DQSGWGPISYANGSGPDQRPYCMWHYPKPCGIVPAKSGVGPYCFTPSPVVGTTDRSGAP	120		
Db	466	dggwgpisyangsgpdqrpwcwhypkpcgivpaksqvgpycftpspvvgttdrsgap	525		
Qy	121	TSWSGENDTDVFLNNTREPLGNWFGCTTMMNSTGCTFKVCGAPPCTVGGAGNNTLHCPTDC	180		

Db	526	tyswgendtdvfvlnntrpplgnwfgctwmnstgftkvcgappcvlgagntnlhcptdc	585
QY	181	FRKHDPATYSCRGSGPWITPRCLVDYYPYRLWHVPCINVTIFKIRMYVGVGVEHRLBAACN	240
Db	586	frkhpdatysrcsgspwiprcldvdpvryrlwhypctinytifkirmvvgvvehrleaacn	645
QY	241	WTRGERCDLEDNR 254	
Db	646	wtrgercdledr 659	
RESULT	11		
R31621			
ID	R31621	standard; protein; 3011 AA.	
XX	R31621;		
XX	24-MAY-1993	(first entry)	
DT	XX	Hepatitis C virus (HCV) polyprotein.	
DE	XX	Hepatitis C virus (HCV) polyprotein.	
XX	XX	Hepatitis; liver disease; HCV1; monoclonal antibody; epitope;	
KW	XX	immobilised reagent; immunoassay; diagnosis; detection; treatment;	
KW	XX	infection.	
OS	XX	Hepatitis C virus type 1.	
XX	XX	Hepatitis C virus type 1.	
FH	Key	Location/Qualifiers	
FT	Domain	1..191	
FT	/label=	"C domain"	
FT	/note=	"nucleocapsid protein"	
FT	Domain	192..383	
FT	/label=	"E1"	
FT	/note=	"virion envelope protein"	
FT	Domain	384..800	
FT	/label=	"E2/NS1"	
FT	/note=	"possible envelope"	
FT	Domain	800..1050	
FT	/label=	"NS2"	
FT	/note=	"unknown function"	
FT	Domain	1050..1650	
FT	/label=	"NS3"	
FT	/note=	"putative protease domain"	
FT	Domain	1651..2100	
FT	/label=	"NS4"	
FT	/note=	"unknown function"	
FT	Domain	2100..3011	
FT	/label=	"NS5"	
FT	/note=	"polymerase"	
XX	XX	WO9300365-A.	
XX	XX	07-JAN-1993.	
XX	XX	24-JUN-1992; 92WO-US05388.	
XX	XX	24-JUN-1991; 91US-0722489.	
XX	XX	(CHIR ) CHIRON CORP.	
XX	XX	Chien DY, Rutter W;	
XX	XX	WPI; 1993-036334/04.	
XX	XX	Polyptide(s) comprising truncated hepatitis C virus sequences -	
XX	XX	for detection, prevention and treatment of hepatitis C infection	
XX	XX	Claim 1; Fig 1; 80pp; English.	
XX	XX	This sequence represents the polyprotein of the HCV prototype	
CC	CC	isolate HCV1. When compared with all known viral sequences, small	
CC	CC	but significant co-linear homologies are observed with the non-	
CC	CC	structural proteins of the flavivirus family, and with the	

CC pestiviruses. The domains shown in the features table are however,  
CC tentatively assigned. The polyprotein, and epitopes of it are  
CC useful for inducing immunological response in a subject against  
CC HCV. The presence of Abs against HCV can be detected using an  
CC immunoassay.  
XX  
SQ Sequence 3011 AA;

Query Match 99.9%; Score 1492; DB 14; Length 3011;  
Best Local Similarity 99.6%; Pred. No. 1.3e-123;  
Matches 253; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
Db 406 gakqnvqlintngswhlNSTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 465  
QY 61 DQGWPISYANGSGDPQRPYCHWHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120  
Db 466 dqgwgplsyangsgdpqrpYCHWHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 525  
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGTFKVCAPPVCVIGGAGNNTLHCPTDC 180  
Db 526 tyswgendtdfvlNTRPPLGNWFGCTWMNSTGTFKVCAPPVCVIGGAGNNTLHCPTDC 585  
QY 181 FRKHDPATYSRCGSGPWINTPRCLVDYPYRLWHYPCTINVTIFKIRMYVGGVHRLEAACN 240  
Db 586 frkhdpatysrcgsgpwintprclvdypYRLWHYPCTINVTIFKIRMYVGGVHRLEAACN 645  
QY 241 WTRGERCDLEDRDR 254  
Db 646 wtrgercdleddr 659

RESULT 12  
R08124  
ID R08124 standard; protein; 2955 AA.  
AC R08124;  
XX  
DT 23-JAN-1991 (first entry)  
XX  
DE Hepatitis C virus putative polyprotein.  
XX  
KW Hepatitis C virus (HCV); antiviral agent.  
XX  
OS Hepatitis C virus.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 9..9 /label-K or R  
FT Misc-difference 11..11 /label-N or T  
FT Misc-difference 176..176 /label-I or T  
FT Misc-difference 334..334 /label-M or V  
FT Misc-difference 603..603 /label-I or L  
FT Misc-difference 848..848 /label-Y or N  
FT Misc-difference 1114..1114 /label-P or S  
FT Misc-difference 1117..1117 /label-S or T  
FT Misc-difference 1276..1276 /label-P or L  
FT Misc-difference 1454..1454 /label-C or Y  
FT Misc-difference 1471..1471 /label-T or S  
FT Misc-difference 1877..1877 /label-E or G

FT Misc-difference 1948..1948 /label-L or H  
FT Misc-difference 1949..1949 /label-S or C  
FT Misc-difference 2021..2021 /label-V or G  
FT Misc-difference 2349..2349 /label-T or S  
FT Misc-difference 2385..2385 /label-Y or F  
FT Misc-difference 2386..2386 /label-S or A  
FT Misc-difference 2502..2502 /label-L or F  
FT Misc-difference 2690..2690 /label-R or G  
FT Misc-difference 2921..2921 /label-R or G  
XX  
PN EP388232-A.  
XX  
PD 19-SEP-1990.  
XX  
PE 16-MAR-1990; 90EP-0302866.  
XX  
PR 18-MAY-1989; 89US-0355002.  
PR 17-MAR-1989; 89US-0325338.  
PR 20-APR-1989; 89US-0341334.  
XX  
PA (CHIR-) CHIRON CORP.  
XX  
PI Houghton M, Choo QL, Kuo G;  
DR WPI; 1990-284418/38.  
DR N-PSDB; Q05956.  
XX  
PT Hepatitis C virus DNA - used for producing probes,  
PT polypeptide(s), antibodies and anti-sense polynucleotide(s) for  
PT diagnosis and therapy.  
XX  
PS Disclosure: Fig 17; 83pp; English.  
XX  
CC HCV cDNA libraries were constructed using pooled serum from a  
CC chimpanzee with chronic HCV infection. A lambda gt10 library was  
CC screened with probes derived from previously isolated clones. The  
CC ORF is derived from the overlapping clones b114a, ag30a, CA205a,  
CC CA290a, CA216a, p14a, CA167b, CA156e, CA84a, CA59a, K9-1, 261, 131,  
CC 12f, 141, 11b, 7f, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c,  
CC 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a and 16jh.  
CC Polypeptide encoded by this sequence can be used to design probes  
CC for the detection of HCV nucleic acids, in screening programmes  
CC for antiviral agents and in preparing blood free of HCV. The  
CC sequence contains 188 (overlapping) peptides which are claimed as  
CC HCV epitopes.  
CC See also Q05955.  
XX  
SQ Sequence 2955 AA;

Query Match 99.7%; Score 1489; DB 11; Length 2955;  
Best Local Similarity 99.6%; Pred. No. 2.3e-123;  
Matches 253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
Db 406 gakqnvqlintngswhlNSTALNCNDLSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 465  
QY 61 DQGWPISYANGSGDPQRPYCHWHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120  
Db 466 dqgwgplsyangsgdpqrpYCHWHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 525  
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGTFKVCAPPVCVIGGAGNNTLHCPTDC 180

Db 526 tyswgeandtvdvlntrppignwfgctwmnstgftkvcgappcvgaggnntlhcptdc 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRLAECN 240  
 |||||  
 Db 586 frkhpdatysrcsgpwtprclvdypylwhypctinytlfclrmvvggvehrleaecn 645  
 |||||

QY 241 WTRGERCDLEDRDR 254  
 |||||  
 Db 646 wtrgercdleldr 659

RESULT 13

R34009  
 ID R34009 standard; Protein; 2816 AA.  
 XX  
 AC R34009;  
 XX  
 DT 26-JUL-1993 (first entry)  
 XX  
 DE HCV-1 polyprotein.  
 DE  
 XX  
 KW Polymerase chain reaction; PCR; amplify; primer; hepatitis C virus;  
 KW HCV; asymptomatic; chronically infected; epitope; viral isolate;  
 KW domain; immunological; cross-reactive; envelope protein; vaccine;  
 KW gp53(BVDV)/gp55; hog cholera virus; pestivirus; NS1; flavivirus.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PN WO9306126-A.  
 XX  
 PD 01-APR-1993.  
 XX  
 PF 11-SEP-1992; 92WO-US07683.  
 XX  
 PR 13-SEP-1991; 91US-0759575.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Houghton M, Weiner AJ;  
 XX  
 DR WPT; 1993-117468/14.  
 XX  
 PT Immuno-reactive hepatitis C virus polypeptide compsns. - contg.  
 PT at least 2 sequences from the first variable domain of distinct  
 PT HCV isolates  
 XX  
 PS Disclosure; Fig 9; 106pp; English.  
 CC  
 CC This sequence represents the entire hepatitis C virus polyprotein.  
 CC HCV is a member of the flavivirus family and appears to encode a basic  
 CC polypeptide domain ("C") at the N-terminal of the viral polyprotein,  
 CC followed by two glycoprotein domains ("E1", "E2/NS1"), upstream of the  
 CC nonstructural genes NS2 through NS5. See also Q39134-48, R33982-  
 CC 4008 and R38088-89.  
 CC  
 XX Sequence 2816 AA;  
 SQ

Query Match 99.5%; Score 1486; DB 14; Length 2816;  
 Best Local Similarity 99.6%; Pred. No. 3.9e-123;  
 Matches 253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60  
 |||||  
 Db 406 gakqnvqlintngswhlNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DQGWGPISYANGSGDPDRPYCWHYPKPCGIVPAKSVCGPVYCTFPTSPVVGTTDRSGAP 120  
 |||||  
 Db 466 dgqwgpiisyangsgdgrpycwhypkpcgivpaksvcgpycftpspvvgttdrsgap 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180  
 |||||  
 Db 526 tyswgeandtvdvlntrppignwfgctwmnstgftkvcgappcvgaggnntlhcptdc 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRLAECN 240  
 |||||  
 Db 586 frkhpdatysrcsgpwtprclvdypylwhypctinytlfclrmvvggvehrleaecn 645  
 |||||

QY 241 WTRGERCDLEDRDR 254  
 |||||  
 Db 646 wtrgercdleldr 659

RESULT 14

R33185  
 ID R33185 standard; peptide; 621 AA.  
 XX  
 AC R33185;  
 XX  
 DT 03-JUL-1993 (first entry)  
 XX  
 DE Sequence of subfragment PHCV80 (AAs 365-731) of the hepatitis  
 DE C virus (HCV) genome.  
 DE  
 XX  
 KW Immunogenic peptide; hepatitis C virus; immunogenic domain;  
 KW monoclonal antibody; diagnosis; detection; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9304205-A.  
 XX  
 PD 04-MAR-1993.  
 XX  
 PF 21-AUG-1992; 92WO-US07189.  
 XX  
 PR 21-AUG-1991; 91US-0748292.  
 XX  
 PA (ABBO ) ABBOTT LAB.  
 XX  
 PI Dailey SH, Desai SM, Devare SG, Johnson JE, Mehta SU;  
 XX  
 DR WPI; 1993-094030/11.  
 XX  
 PT Monoclonal antibodies specific for hepatitis C virus E2-NS1  
 PT antigen - useful for diagnosis and evaluation of HCV infections  
 PT and in differentiation studies  
 XX  
 PS Example; Pages 31-34; 48pp; English.  
 CC  
 CC Immunogenic domains of E2/NS1 region of HCV genome encompassing AAs  
 CC 600-720 were mapped with PEPSCAN analysis. Based on the EIA reactivity  
 CC of a panel of HCV positive sera, peptide R33184 was chosen as the  
 CC immunogen for the generation of monoclonal antibodies to HCV NS1.  
 CC Several individual oligos representing AAs 365-731 of HCV genome  
 CC were ligated and cloned as three separate EcoRI/BamHI subfragments  
 CC into the CKS fusion vector pJ0200. The sequences of these three  
 CC subfragments - PHCV80, PHCV77 and PHCV65 are in R33185-7.  
 CC Analysis showed that MAbS H13C113 and H23C163 showed reactivity  
 CC PHCV80 and PHCV65.  
 CC  
 XX Sequence 621 AA;  
 SQ

Query Match 98.6%; Score 1473; DB 14; Length 621;  
 Best Local Similarity 97.3%; Pred. No. 9.5e-123;  
 Matches 253; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60  
 |||||  
 Db 290 gakqnvqlintngswhlNSTALNCNDSINTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 349

QY 61 DQGWGPISYANGSGDPDRPYCWHYPKPCGIVPAKSVCGPVYCTFPTSPVVGTTDRSGAP 120  
 |||||  
 Db 350 dgqwgpiisyangsgdgrpycwhypkpcgivpaksvcgpycftpspvvgttdrsgap 409

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGA-----PPCVIGAGNNTL 174



Db 410 tyswgentdvfvlntnrplgnwfgctwmnstgftkvcgappcvlgppcvigagnttl 469  
Qy 175 HCPDTCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCTINYTFIKIRMYGVGEHR 234  
Db 470 hcptdcfrkhpdatysrcsgpwtprclvdypylwhypctinytifikirmyvggvehr 529  
Qy 235 LEACNWTGRGCDLEDRDR 254  
Db 530 leaecnwtgrgcdledrdr 549

RESULT 15

R33591  
ID R33591 standard; protein; 622 AA.  
XX  
AC R33591;  
XX  
DT 05-JUL-1993 (first entry)  
XX  
DE HCV CKS-NS1S1-NS1S2 fusion antigen.  
XX  
DE Hepatitis C Virus; non-A, non-B hepatitis virus; NANBH;  
KW non-structural protein; CMP-KDO synthetase; CKS fusion protein;  
KW CTP:CMP-3-deoxy-manno-octulosonate cytidyl transferase;  
KW immunoassay; pHCV-80.  
XX  
OS Hepatitis C Virus.

XX Key Location/Qualifiers  
FH Misc-difference 622  
FT /note= "not defined"

FT  
PN W09304088-A.  
XX  
XX  
PD 04-MAR-1993.  
XX  
XX 21-AUG-1992; 92WO-US07188.  
XX  
XX 21-AUG-1991; 91US-0748561.  
XX  
XX (ABBO ) ABBOTT LAB.

XX Dailey SH, Desai SM, Devare SG;  
XX WPI; 1993-093941/11.  
XX

XX Hepatitis C assay using recombinant NS1 region antigens - for  
PT detecting antibodies and antigen in body fluids from individuals  
PT exposed to hepatitis C virus

XX Claim 4; Page 44-46; 175pp; English.

XX The construction of pHCV-80 (NS1S1-NS1S2) involved using the  
CC SacI/BamHI insert from pHCV-65 (see R33589) and ligating that into  
CC the SacI/BamHI vector backbone of pHCV-77 (see R33588). The resultant  
CC HCV gene represents amino acids 365-731 of the HCV genome. This  
CC resulted in a 1101bp EcoRI/BamHI fragment of HCV cloned into the  
CC CKS fusion vector pJO200. The complete amino acid sequence of the  
CC antigen is designated pHCV-80 (i.e. R33591). The resultant fusion  
CC protein HCV CKS NS1S1-NS1S2 consists of 239 amino acids of CKS,  
CC seven amino acids contributed by linker DNA sequences and 367 amino  
CC acids from the NS1 region of the HCV genome. The fusion protein is  
CC used to detect antibodies and antigen in body fluids from  
CC individuals exposed to HCV.

XX Sequence 622 AA;

Query Match 98.6%; Score 1473; DB 14; Length 622;  
Best Local Similarity 97.3%; Pred. No. 9.5e-123;  
Matches 253; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

Qy 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLAGLFYHHKFNSSGCPERLASCRPI.TDF 60  
Db 290 gakgnvqlintngswhlntalnncndsintgwlaglffyhhkfnssgcpelascrpltdf 349  
Qy 61 DOGWGPTISYANGSGPQORPYCWHYPPKPCGIVPAKSVCGPVYCFTPSPVVGTTDRSGAP 120  
Db 350 dqgwqglsyangsgpqrpycwhypkpcgvlvpaksvcgvpycftpspvvvgtdrsgap 409  
Qy 121 TYSWGENDDTVFVLNNTRPPLGNMFGCTWMNSTGFTKVCGA-----PPCVIGGAGNNTL 174  
Db 410 tyswgentdvfvlntnrplgnwfgctwmnstgftkvcgappcvlgppcvigagnttl 469  
Qy 175 HCPDTCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCTINYTFIKIRMYGVGEHR 234  
Db 470 hcptdcfrkhpdatysrcsgpwtprclvdypylwhypctinytifikirmyvggvehr 529  
Qy 235 LEACNWTGRGCDLEDRDR 254  
Db 530 leaecnwtgrgcdledrdr 549

Search completed: March 6, 2001, 11:51:51  
Job time: 74 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:15 ; Search time 45.68 Seconds  
(without alignments)  
377.556 Million cell updates/sec

Title: US-09-407-430-3  
Perfect score: 1494  
Sequence: 1 GAKQNVQLINTNGSWHLNST.....LEAACNWTGRGCDEDRD 254

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched: 195891 seqs, 67900655 residues
Total number of hits satisfying chosen parameters: 195891
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %	Score	Match	Length	DB	ID	Description
1	1494	100.0	3011	1	GNWVC3	genome polyprotein	
2	1438	96.3	3011	1	S40770	genome polyprotein	
3	1398	93.6	3011	1	GNWVC	genome polyprotein	
4	1367	91.5	716	2	JQ1366	polyprotein - hepa	
5	1273	85.2	3010	1	GNWVW	genome polyprotein	
6	1272	85.1	640	2	JQ1584	genome polyprotein	
7	1254	83.9	3010	1	GNWVC	genome polyprotein	
8	1253	83.9	3010	1	GNWVCJ	genome polyprotein	
9	1244	83.3	782	2	J19875	genome polyprotein	
10	1242.5	83.2	3014	1	JC5620	genome polyprotein	
11	1236	82.7	782	2	S18876	genome polyprotein	
12	1234	82.6	3010	1	A45573	genome polyprotein	
13	1234	82.6	3010	1	S18030	genome polyprotein	
14	1228	82.2	782	2	S18032	genome polyprotein	
15	1199	80.3	782	2	S18031	genome polyprotein	
16	1171	78.4	3033	2	PN0677	genome polyprotein	
17	1170	78.3	3033	1	GNWV38	hypothetical prote	
18	1156	77.4	3033	1	JQ1303	genome polyprotein	
19	1122	75.1	350	2	S35631	genome polyprotein	
20	942	63.1	234	2	S32742	genome polyprotein	
21	936	62.7	235	2	S32747	genome polyprotein	
22	905	60.6	237	2	S32744	genome polyprotein	
23	756	50.6	415	2	PC4407	genome polyprotein	
24	649.5	43.5	876	2	PC2219	envelope protein -	
25	597	40.0	513	2	A44150	polypeptide - hepa	
26	550	36.8	138	2	S24080	structural protein	
27	539	36.1	138	2	S24081	envelope protein -	
28	522	34.9	523	2	JQ1926	polyprotein - hepa	
29	521	34.9	138	2	S24074	envelope protein -	

## ALIGNMENTS

RESULT 1

genome polyprotein - hepatitis C virus (strain HCV-1)  
N:Contains: capsid protein C; envelope protein M; envelope  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 26-May-2000  
C:Accession: A39166; PQ0404  
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;  
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A:Title: Genetic organization and diversity of the hepatitis C virus.  
A:Reference number: A39166; MUID:91172826  
A:Accession: A39166  
A:Molecule type: mRNA  
A:Residues: 1-3011 <CHO>  
A:Cross-references: GB:M62321; NID:9329873; PIDN:AAA45676.1; PID:g329874  
R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap,  
J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship t  
A:Reference number: PQ0393; MUID:92268871

```
Query Match      100.0%; Score 1494; DB 1; Length 3011;
Best Local Similarity 100.0%; Pred. NO. 1.5e-113;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
|||||
Db 406 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DQGWGPIYANGSGPDQRPYCWHPYPPKPGIVPAKSVCGVPVYCFPTSPVVVGTTRDSGAP 120
|||||
Db 466 DQGWGPIYANGSGPDQRPYCWHPYPPKPGIVPAKSVCGVPVYCFPTSPVVVGTTRDSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPCCVIGGAGNNTLHCPTDC 180
|||||
Db 526 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPCCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPDATYSRCGSPWITPCLVDYPYRLWHYPCTINTYIFKIRMYVGGVEHRLAECN 240
|||||
Db 586 FRKHDPDATYSRCGSPWITPCLVDYPYRLWHYPCTINTYIFKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254
|||||
Db 646 WTRGERCDLEDRDR 659

RESULT 2
S40770
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
A:Accession: S40770; PC1285
R:Okamoto, H.
submitted to the EMBL Data Library, March 1992
A:Reference number: S40770
A:Accession: S40770
A:Molecule type: genomic RNA
A:Residues: 1-3011 <OKA>
A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda,
Jpn. J. Exp. Med. 60, 167-177, 1990
A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
A:Reference number: PC1284; MUID:91013116
A:Accession: PC1285
A:Molecule type: genomic RNA
A:Residues: 1-513 <OK2>
A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
A:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transm
F:116-191/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 96.3%; Score 1438; DB 1; Length 3011;
Best Local Similarity 94.9%; Pred. No. 5e-109;
Matches 241; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
|||||
Db 406 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DQGWGPIYANGSGPDQRPYCWHPYPPKPGIVPAKSVCGVPVYCFPTSPVVVGTTRDSGAP 120
|||||
Db 466 DQGWGPIYANGSGPDQRPYCWHPYPPKPGIVPAKSVCGVPVYCFPTSPVVVGTTRDSGAP 525
```

```
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPCCVIGGAGNNTLHCPTDC 180
|||||
Db 526 TYNKANDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPCCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPDATYSRCGSPWITPCLVDYPYRLWHYPCTINTYIFKIRMYVGGVEHRLAECN 240
|||||
Db 586 FRKHDPDATYSRCGSPWITPCLVDYPYRLWHYPCTINTYIFKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254
|||||
Db 646 WTRGERCDLEDRDR 659

RESULT 3
GNVWCH
genome polyprotein - hepatitis C virus (strain H)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 26-May-2000
C:Accession: A36814; A41546
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
submitted to GenBank, July 1992
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus
A:Reference number: A36814
A:Accession: A36814
A:Molecule type: genomic RNA
A:Residues: 1-3011 <INC>
A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA5534.1; PID:g329738
R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: comp
A:Reference number: A41546; MUID:92052256
A:Contents: annotation
A:Note: neither amino acid nor nucleotide sequence is given
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>
F:1312-1317/Region: nucleotide-binding motif A (P-loop)
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 93.6%; Score 1398; DB 1; Length 3011;
Best Local Similarity 93.3%; Pred. No. 1e-105;
Matches 237; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
|||||
Db 406 GAKQNVQLINTNGSWHLNSTALNCNDLSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DQGWGPIYANGSGPDQRPYCWHPYPPKPGIVPAKSVCGVPVYCFPTSPVVVGTTRDSGAP 120
|||||
Db 466 DQGWGPIYANGSGPDQRPYCWHPYPPKPGIVPAKSVCGVPVYCFPTSPVVVGTTRDSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPCCVIGGAGNNTLHCPTDC 180
|||||
Db 526 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPCCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPDATYSRCGSPWITPCLVDYPYRLWHYPCTINTYIFKIRMYVGGVEHRLAECN 240
|||||
```

Db 586 FRKYPEATYSRGSGPRITPRCMVDYPRYLWHYPCTINTTIFKVRMYVGVGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254

Db 646 WTRGERCDLEDRDR 659

# RESULT 4

JQ1366

polyprotein - hepatitis C virus (French isolate) (fragments)

C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 29-May-1998

C:Accession: JQ1366

J:Krensdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

R:Gen. Virol. 72, 2557-2561, 1991

A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A:Reference number: JQ1366; MUID:92013977

A:Accession: JQ1366

A:Molecule type: genomic RNA

A:Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: glycoprotein; polyprotein

F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

# Query Match

Best Local Similarity 91.5%; Score 1367; DB 2; Length 716;

Matches 222; Conservative 25; Mismatches 74; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60

Db 73 GAKQDIQLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLADP 132

QY 61 DQGWGPISYANGSGPDQRPYCHYPPKPGCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

Db 133 DQGWGPISYANGSGPDQRPYCHYPPKPGCGIVPAQVCGPVYCFPTSPVVGTTKLGAP 192

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180

Db 193 TTNMGENDTDFVLNTRPPLGNWFGCTWMNSGFTKVCAPPVCVIGGAGNNTLHCPTDC 252

QY 181 FRKHDPATYSRSGSGPWITPRCLVDYPRYLWHYPCTINTTIFKIRMYVGVGVEHRLAECN 240

Db 253 FRKHPEATYSRSGSGPWITPRCLVGYPRYLWHYPCTINTTIFKIRMYVGVGVEHRLQVACN 312

QY 241 WTRGERCDLEDRDR 254

Db 313 WTRGERCNLEDRDR 326

# RESULT 5

GNVTV

genome polyprotein - hepatitis C virus (strain Taiwan)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 26-May-2000

C:Accession: A40244

R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.

Virology 188, 102-113, 1992

A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the

A:Reference number: A40244; MUID:92230206

A:Accession: A40244

A:Molecule type: genomic RNA

A:Residues: 1-3010 <CHE>

A:Cross-references: GB:M84754

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural

F:1-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <ME>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1815/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology <NS3>

F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>

F:1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,204,

# Query Match

Best Local Similarity 85.2%; Score 1273; DB 1; Length 3010;

Matches 208; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60

Db 406 GASQKIQILINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERMASCRSDIKF 465

QY 61 DQGWGPISYANGSGPDQRPYCHYPPKPGCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

Db 466 DQGWGPITVTEADIQDQRPYCHYPPKPGCGIVPAQVCGPVYCFPTSPVVGTTDRFGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180

Db 526 TYSWGENETDFVLNTRPPLGNWFGCTWMNSTGFTKTCGGPPCNTGGGNNNTLVCPDTC 585

QY 181 FRKHDPATYSRSGSGPWITPRCLVDYPRYLWHYPCTINTTIFKIRMYVGVGVEHRLAECN 240

Db 586 FRKHPEATYTKCGSPWLPTRCMVDYPRYLWHYPCTINTTIFKIRMYVGVGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254

Db 646 WTRGERCDLEDRDR 659

# RESULT 6

JQ1584

genome polyprotein - hepatitis C virus (strain U.K.) (fragment)

N:Contains: core protein C; envelope protein E1; envelope protein E2; nonstructural

C:Species: hepatitis C virus

C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 28-May-1999

C:Accession: JQ1584

R:Kumar, U.; Cheng, D.; Thomas, H.; Monjardino, J.

J. Gen. Virol. 73, 1521-1525, 1992

A:Title: Cloning and sequencing of the structural region and expression of putative c

A:Reference number: JQ1584; MUID:92300349

A:Accession: JQ1584

A:Molecule type: genomic RNA

A:Residues: 1-640 <KUM>

A:Cross-references: GB:X84079; NID:G643119; PIDN:CAA58888.1; PID:G643120

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: core protein; envelope protein; glycoprotein; nonstructural protein; poly

F:1-151/Product: core protein C #status predicted <CPC>

F:192-389/Product: envelope protein E1 #status predicted <EEL>

F:390-640/Product: envelope protein E2 and nonstructural protein NS1 #status predite

F:196,209,234,305,417,430,448,476,540,556,576,623/Binding site: carbohydrate (Asn) (c

# Query Match

Best Local Similarity 85.1%; Score 1272; DB 2; Length 640;

Matches 208; Conservative 17; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60

Db 406 GSRQNLQILINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCTERMASCRPLADP 465

QY 61 DQGWGPISYANGSGPDQRPYCHYPPKPGCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120

Db 466 DQGWGPITVTEADIQDQRPYCHYPPKPGCGIVPAQVCGPVYCFPTSPVVGTTDRSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCVIGGAGNNTLHCPTDC 180

Db 526 TYNWGSNDTDVFLVNTNRPPLGNFGCTWNNSSGFTKVCAGPCCNIGGVGNNTLHCPTDC 585  
QY 181 FRKHDPATYSCGSGPWTTPCLVDYDYLRLWHYPCTTNTYTFKIRMYVGVGVEHRL 235  
Db 586 FRKHPEATYSCGSGPWTTPCLVDYDYLRLWHYPCTTNTYTFKIRMYVGVGVEHRL 640

## RESULT 7

GNWVTC

genome polyprotein - hepatitis C virus  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 26-May-2000  
C:Accession: A38465  
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
A:Title: Structure and organization of the hepatitis C virus genome isolated from human  
A:Reference number: A38465; MUID:91140698  
A:Accession: A38465  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <AK>  
A:Cross-references: EMBL:M58335; MID:g329770; PIDN:AAAT2945.1; PID:g329771  
A:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS2 #status predicted <NS2>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 83.9%; Score 1254; DB 1; Length 3010;  
Best Local Similarity 81.1%; Pred. No. 5.3e-94;  
Matches 206; Conservative 16; Mismatches 32; Indels 0; Gaps 0;

QY 1 GAKQVQLINTNGSWHLNLTALNCNDSLTGWLGLFHHKFNSSGCGPERLASCRPLTDF 60  
Db 406 GPSOKIQLINTNGSWHLNLTALNCNDSLTGFLAALFYTHSFNSGCGPERMAQCRTIDKF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSVCGVPYCTPSPVVGTTDRSGAP 120  
Db 466 DOGNGPITYAESRSDQRPYCHWYPPQCTIVPASEVCGVPYCTPSPVVGTTDRGVP 525

QY 121 TYSNGENDTDVFLNTRPPLGNFGCTWNNSSGFTKVCAGPCCNIGGVGNNTLHCPTDC 180  
Db 526 TYRGENETDVLNTRPPOGNNFGCTWNNSSGFTKTCGGPPCNGGVGNNTLTCPTDC 585

QY 181 FRKHDPATYSCGSGPWTTPCLVDYDYLRLWHYPCTTNTYTFKIRMYVGVGVEHRLAECN 240  
Db 586 FRKHPEATYSCGSGPWTTPCLVDYDYLRLWHYPCTTNTYTFKIRMYVGVGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254  
Db 646 WTRGERCDLEDRDR 659

## RESULT 8

GNWVCJ

genome polyprotein - hepatitis C virus (strain J)  
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Jun-2000  
C:Accession: A39253; PS0086  
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shi  
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patient  
A:Reference number: A39253; MUID:91088550  
A:Accession: A39253  
A:Molecule type: genomic RNA  
A:Residues: 1-3010 <KAT>  
A:Cross-references: GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611  
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence v  
A:Reference number: PS0085  
A:Accession: PS0086  
A:Molecule type: genomic RNA  
A:Residues: 2650-2707 <KA2>  
A:Experimental source: Japanese isolate  
C:Comment: The cleavage sites of this polyprotein have not been determined.  
C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; tra  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
F:1230-1493/Domain: DEAD/H box helicase homology <DEAD>  
F:1312-1317/Region: nucleotide-binding motif A (P-loop)  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,224

Query Match 83.9%; Score 1253; DB 1; Length 3010;  
Best Local Similarity 79.5%; Pred. No. 6.4e-94;  
Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQVQLINTNGSWHLNLTALNCNDSLTGWLGLFHHKFNSSGCGPERLASCRPLTDF 60  
Db 406 GPSOKIQLINTNGSWHLNLTALNCNDSLTGFLAALFYTHSFNSGCGPERMASCRPIDEF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSVCGVPYCTPSPVVGTTDRSGAP 120  
Db 466 AQGNGPITHDPESDQRPYCHWYAPPCGIVPAQVCGVPYCTPSPVVGTTDRFGAP 525

QY 121 TYSNGENDTDVFLNTRPPLGNFGCTWNNSSGFTKVCAGPCCNIGGVGNNTLHCPTDC 180  
Db 526 TYSNGENETDVLNTRPPOGNNFGCTWNNSSGFTKTCGGPPCNGGVGNNTLVCPTDC 585

QY 181 FRKHDPATYSCGSGPWTTPCLVDYDYLRLWHYPCTTNTYTFKIRMYVGVGVEHRLAECN 240  
Db 586 FRKHPEATYSCGSGPWTTPCLVDYDYLRLWHYPCTTNTYTFKIRMYVGVGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254  
Db 646 WTRGERCDLEDRDR 659

## RESULT 9

S19875

genome polyprotein - hepatitis C virus (isolate JK3) (fragment)  
N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein  
C:Species: hepatitis C virus  
A:Variety: isolate JK3  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 26-Aug-1999  
C:Accession: S19875  
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
Submitted to the EMBL Data Library, September 1991  
A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus

A:Reference number: S18029

A:Accession: S19875

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:CROSS-references: EMBL:X61592; NID:G59482; PIDN:CAA43789.1; PID:G59483

A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: core protein #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 83.3%; Score 1244; DB 2; Length 782;

Best Local Similarity 79.5%; Pred. No. 8.6e-94;

Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQVQLINTNGSHLNSLTALNCNDSLTGWLGLFHHKFNSSGCGPERLASCRLPTDF 60

DB 406 GPSONIQVLVNSGSHINRTALSCNDSLTGFLAALFYTHKFENASGCGPERWASCRSDITF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSGVCGPVYCFPTSPVVVGTDRSGAP 120

DB 466 DOGNGPITHVPNTDQKPYCHWYAPRPGCGIVPAQVCGPVYCFPTSPVVVGTDRSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 180

DB 526 TTTWGENETDVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 585

QY 181 FRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 240

DB 586 FRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254

DB 646 WTRGERCDLEDRDR 659

RESULT 10

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000

A:Accession: JC5620

R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.

Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A:Reference number: JC5620; MUID:97366593

A:Accession: JC5620

A:Molecule type: mRNA

A:Residues: 1-3014 <CHA>

A:CROSS-references: GB:Y13184

A:Experimental source: genotype 5a, which predominates in South Africa

A:Note: the translation of the nucleotide sequence is not complete in this paper

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <EPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:384-408/Region: hypervariable #status predicted

F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>

F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>

F:1008-1616/Product: hepatitis C virus genome polyprotein; serine proteinase; transmembrane

F:1231-1494/Domain: DEAD/H box helicase homology <DEAD>

F:1231-1238/Region: nucleotide-binding motif A (P-loop)

F:1313-1318/Region: nucleotide-binding motif B

F:1317-1320/Region: DEXH motif

F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>

F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>

F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>

F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 83.2%; Score 1242.5; DB 1; Length 3014;

Best Local Similarity 78.8%; Pred. No. 4.6e-93;

Matches 201; Conservative 26; Mismatches 27; Indels 1; Gaps 1;

QY 1 GAKQVQLINTNGSHLNSLTALNCNDSLTGWLGLFHHKFNSSGCGPERLASCRLPTDF 60

DB 406 GPQQLQFVNTNGSHLNSLTALNCNDSLTGWLGLFHHKFNSSGCGPERLASCRLPTDF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSGVCGPVYCFPTSPVVVGTDRSGA 119

DB 466 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSGVCGPVYCFPTSPVVVGTDRSGA 525

QY 120 PTYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 179

DB 526 PTYNWSNETDILLNTRPAGNWFCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 585

QY 180 CFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 239

DB 586 CFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 645

QY 240 NWTGERCDLEDRDR 254

DB 646 NWTGERCDLEDRDR 660

RESULT 11

S19876

genome polyprotein - hepatitis C virus (isolate JK5) (fragment)

N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein

C:Species: hepatitis C virus

A:Variety: isolate JK5

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 26-Aug-1999

C:Accession: S19876

R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus

A:Reference number: S18029

A:Accession: S19876

A:Molecule type: genomic RNA

A:Residues: 1-782 <HON>

A:CROSS-references: EMBL:X61595; NID:G59486; PIDN:CAA43792.1; PID:G59487

A:Experimental source: isolate JK5

C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology

C:Keywords: capsid protein; core protein; envelope protein; glycoprotein; nonstructural

F:1-191/Product: core protein #status predicted <MAT1>

F:192-383/Product: core protein #status predicted <MAT2>

F:384-733/Product: NS1/E2 protein #status predicted <MAT3>

F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 82.7%; Score 1236; DB 2; Length 782;

Best Local Similarity 79.5%; Pred. No. 3.9e-93;

Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQVQLINTNGSHLNSLTALNCNDSLTGWLGLFHHKFNSSGCGPERLASCRLPTDF 60

DB 406 GPSONIQVLVNSGSHINRTALSCNDSLTGFLAALFYTHKFENASGCGPERWASCRSDITF 465

QY 61 DOGNGPISYANGSGDQRPYCHWYPPKPGIVPAKSGVCGPVYCFPTSPVVVGTDRSGAP 120

DB 466 DOGNGPITHVPNTDQKPYCHWYAPRPGCGIVPAQVCGPVYCFPTSPVVVGTDRSGAP 525

QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 180

DB 526 TYNWGANETDVLNTRPPOGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNTLHCPTDC 585

QY 181 FRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 240

DB 586 FRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 645





C:Accession: S18032  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso  
 A:Reference number: S18029  
 A:Accession: S18032  
 A:Molecule type: genomic RNA  
 A:Residues: 1-782 <HON>  
 A:Cross-references: EMBL:X61594  
 A:Experimental source: Isolate JK4  
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
 F:1-191/Product: capsid protein; core protein; envelope protein; glycoprotein; nonstructural  
 F:192-383/Product: core protein #status predicted <MAT1>  
 F:384-733/Product: envelope protein 1 #status predicted <MAT2>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT3>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 82.2%; Score 1228; DB 2; Length 782;  
 Best Local Similarity 79.1%; Pred. No. 1.7e-92;  
 Matches 201; Conservative 23; Mismatches 30; Indels 0; Gaps 0;

QY	1	GAKQNVOLINTNGSHLNSLTALNCNDSLNTGWLGLFHHKFNSSGCPERLASCRPLTDF	60
Db	406	GSQKIQLVNNGSRHINRTALSCNDSLKTGFLAALFYTHKFNAAGCPERMASCCSIDTF	465
QY	61	DQWGPITSYANGSGDQRPYCHYPPKPCGIVPAKSVCGPVYCTPSPVVGTTDRSGAP	120
Db	466	AQWGPITHAESRSSDQRPYCHYAPQPGIVPALQVCGPVYCTPSPVVGTTDRFGAP	525
QY	121	TSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAGAPPVCVIGGAGNNTLHCPTDC	180
Db	526	TYNKGANETDVLNTRPQGTWFGCTWMNSTGFTKTCGGPPCNIGGVGNNTLTCTPTDC	585
QY	181	FRKHDPATYSRCGSPWITPRCLVDYPYRLWHYPCTINTYTIKIRMYVGGVEHRLAAACN	240
Db	586	FRKHPEATYTKCGSPWLTPRCMVHYPYRLWHYPCTVNETVFKVRMYVGGVEHRLTAACN	645
QY	241	WTRGERCDLEDRDR	254
Db	646	WTRGERCNLEDRDR	659

Search completed: March 6, 2001, 11:53:21  
 Job time: 158 sec

## RESULT 15

S18031  
 genome polyprotein - hepatitis C virus (isolate JK2) (fragment)  
 N:Contains: core protein; envelope protein 1; nonstructural protein 2; NS1/E2 protein  
 C:Species: hepatitis C virus  
 A:Variety: isolate JK2  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 29-May-1998  
 C:Accession: S18031  
 R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 submitted to the EMBL Data Library, September 1991  
 A:Description: Sequence analysis of putative structural regions of Hepatitis C Virus iso  
 A:Reference number: S18029  
 A:Accession: S18031  
 A:Molecule type: genomic RNA  
 A:Residues: 1-782 <HON>  
 A:Cross-references: EMBL:X61593  
 A:Experimental source: Isolate JK2  
 C:Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology  
 F:1-191/Product: capsid protein; core protein; envelope protein; glycoprotein; nonstructural  
 F:192-383/Product: core protein #status predicted <MAT1>  
 F:384-733/Product: envelope protein 1 #status predicted <MAT2>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT3>  
 F:734-782/Product: nonstructural protein 2 (fragment) #status predicted <MAT4>

Query Match 80.3%; Score 1199; DB 2; Length 782;  
 Best Local Similarity 76.8%; Pred. No. 4e-90;  
 Matches 195; Conservative 27; Mismatches 32; Indels 0; Gaps 0;

QY	1	GAKQNVOLINTNGSHLNSLTALNCNDSLNTGWLGLFHHKFNSSGCPERLASCRPLTDF	60
----	---	--	----

us-09-407-430-3.rpr

Tue Mar 6 12:02:35 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:53:43 ; Search time 25.34 Seconds  
(without alignments)  
323.706 Million cell updates/sec

Title: US-09-407-430-3  
Perfect score: 1494  
Sequence: 1 GAKQNVQLINTSGWHLNST.....LEACNWTGRCDDLEDRDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1494	100.0	3011	1	POLG_HCV1
2	1398	93.6	3011	1	POLG_HCVH
3	1273	85.2	3010	1	POLG_HCVTW
4	1254	83.9	3010	1	POLG_HCVBK
5	1253	83.9	3010	1	POLG_HCVJA
6	1234	82.6	3010	1	POLG_HCVJT
7	1170	78.3	3033	1	POLG_HCVJ8
8	1158	77.5	737	1	POLG_HCVJ7
9	1156	77.4	3033	1	POLG_HCVJ6
10	1147	76.8	737	1	POLG_HCVJ5
11	511	34.2	520	1	POLG_HCVH4
12	511	34.2	520	1	POLG_HCVH3
13	475	31.8	513	1	POLG_HCVJ2
14	173	11.6	321	1	POLG_HCVTH
15	171	11.4	138	1	POLG_HCVHEO
16	167	11.2	321	1	POLG_HCVH8
17	153	10.2	309	1	POLG_HCVH7
18	110.5	7.4	1700	1	BAR3_CHITE
19	107	7.2	1046	1	PSTA_DICDI
20	103	6.9	1172	1	TSP2_MOUSE
21	97.5	6.5	640	1	UROM_HUMAN
22	96.5	6.5	752	1	CO2_HUMAN
23	94	6.3	555	1	DP87_DICDI
24	93.5	6.3	1172	1	TSP2_HUMAN
25	91.5	6.1	571	1	HENA_P12H
26	91.5	6.1	571	1	HENA_P12HT
27	91	6.1	2871	1	FBNI_HUMAN
28	90	6.0	1170	1	TSP2_BOVIN
29	90	6.0	2871	1	FBNI_MOUSE
30	88.5	5.9	784	1	YAV2_XANCV
31	88.5	5.9	1077	1	SM5A_MOUSE
32	87	5.8	1178	1	TSP1_CHICK
33	87	5.8	2871	1	FBNI_BOVIN

#### RESULT 1

ID	POLG_HCV1	STANDARD	PRT	3011 AA.
AC	P26664			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22); ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21) (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-); NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27); NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66) (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].			
OS	Hepatitis C virus (isolate 1) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-91172826; PubMed-1848704;			
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Welner A.J., Bradley D.W., Kuo G., Houghton M.; "Genetic organization and diversity of the hepatitis C virus." Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).			
RL	CC - FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.			
CC	CC - SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.			
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	CC EMBL; M62321; AAA45676.1; -			
CC	CC PIR; A39166; GNWVC3.			
DR	DR HSSP; P27958; 1HEI.			
DR	DR MEROPS; S29.001; -			
DR	DR MEROPS; U39.001; -			
DR	DR INTERPRO; IPR000745; -			
DR	DR INTERPRO; IPR001490; -			
DR	DR INTERPRO; IPR002166; -			
DR	DR INTERPRO; IPR002518; -			
DR	DR INTERPRO; IPR002519; -			
DR	DR INTERPRO; IPR002521; -			
DR	DR INTERPRO; IPR002522; -			
DR	DR INTERPRO; IPR002531; -			
DR	DR INTERPRO; IPR002868; -			

Q24400 drosophila  
P05160 homo sapien  
P31747 bacillus sp  
Q27969 bos taurus  
P48733 bos taurus  
Q39692 daucus caro  
P12625 alcaligenes  
P56680 lumicola in  
O61483 us musculu  
P75750 scherichia  
P07996 homo sapien  
P22549 dictyostell

#### ALIGNMENTS

34 86.5 5.8 495 1 MLP2\_DROME  
35 85.5 5.7 661 1 F13B\_HUMAN  
36 85.5 5.7 718 1 CDGT\_BACSS  
37 84.5 5.7 532 1 AD50\_BOVIN  
38 84.5 5.7 543 1 UROM\_BOVIN  
39 82.5 5.5 592 1 INV2\_DAUCA  
40 82 5.5 488 1 PHB\_ALCFA  
41 81.5 5.5 402 1 GUN1\_HUMIN  
42 81.5 5.5 722 1 DLL1\_MOUSE  
43 81.5 5.5 818 1 YBGO\_ECOLI  
44 81.5 5.5 1170 1 TSP1\_DICDI  
45 81 5.4 237 1 IPDE\_HUMAN

DR PFAM; PF01560; HCV\_N51; 1.  
 DR PFAM; PF01538; HCV\_N52; 1.  
 DR PFAM; PF01006; HCV\_N54a; 1.  
 DR PFAM; PF01001; HCV\_N54b; 1.  
 DR PFAM; PF01506; HCV\_N55a; 1.  
 DR PFAM; PF00998; HCV\_RDRP; 1.  
 DR PFAM; PF01543; HCV\_capsid; 1.  
 DR PFAM; PF01542; HCV\_Core; 1.  
 DR PFAM; PF01539; HCV\_env; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 100.08; Score 1494; DB 1; Length 3011;  
 Best Local Similarity 100.08; Pred. No. 7.8e-121;  
 Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTGSHLNTALNCNDLSLNTGLAGLFFHHKFNSSGCPERLASCRPLTDF 60  
 DB 406 GAKQNVQLINTGSHLNTALNCNDLSLNTGLAGLFFHHKFNSSGCPERLASCRPLTDF 465  
 QY 61 DQGWGPISYANGSGPDQRYCHYPPKPCGIVPAKSVCGPVYCFPPSPVVGTTDRSGAP 120  
 DB 466 DQGWGPISYANGSGPDQRYCHYPPKPCGIVPAKSVCGPVYCFPPSPVVGTTDRSGAP 525  
 QY 121 TYSWGENDDTVFLNTRPLGNWFCTWNSTGTFKVCAGPCVIGGAGNNTLHCPTDC 180  
 DB 526 TYSWGENDDTVFLNTRPLGNWFCTWNSTGTFKVCAGPCVIGGAGNNTLHCPTDC 585  
 QY 181 FRKHPDATSRCSGSGPWITPRCLVDYPIRLWHYPCTINTYIFKIRMYGVGVEHRLAEACN 240  
 DB 586 FRKHPDATSRCSGSGPWITPRCLVDYPIRLWHYPCTINTYIFKIRMYGVGVEHRLAEACN 645  
 QY 241 WTRGERCDLEDRDR 254

DB 646 WTRGERCDLEDRDR 659

# RESULT 2

POLG\_HCVH STANDARD; PRT; 3011 AA.  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)  
 DE (EC 3.4.22.-); PROTEINASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 RN Hepacivirus.  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE-92052256; PubMed-1658800;  
 RA Inchauspe G., Zebede S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RA "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).  
 RN [2]  
 RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RP MEDLINE-97313122; PubMed-9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RA "Structure of the hepatitis C virus RNA helicase domain.";  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RP MEDLINE-98154321; PubMed-9493270;  
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RA "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode  
 of unwinding.";  
 RL Structure 6:89-100(1998).  
 RL Structure 6:89-100(1998).  
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF  
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE  
 CC ACTIVATION OF NS3.  
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN  
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1  
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.  
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY  
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.  
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; M67463; AAA45534.1;  
 DR PIR; A36814; GNMVCH.  
 DR PDB; 1HEI; 25-NOV-98.  
 DR PDB; 1A1V; 16-FEB-99.  
 DR MEROPS; S29.001; -.



RESULT 4  
POLG\_HCVBK

[illegible]

```

Db      526 TYRGENTDVLNLTNTRPPQGNWFGCGTMMNSGTGFTKTCGGPPCNGVNTLTCTPTDC 585
QY      181 FRKHPDATYSRCGSGPWITTPCLVDYPYRLWHYPCTINVTIFKIRMYGCVGVEHRLBAACN 210
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      586 FRKHPEATYTKCGSGPWLTPRCMVDYPYRLWHYPCTVNFTEFKVRMYGCVGVEHRLNAACN 645
QY      241 WTRGERCDEDRDR 254
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      646 WTRGERCDEDRDR 659
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
POLG_HCVJA
ID      POLG_HCVJA      STANDARD;      PRT;      3010 AA.
AC      P28662;
DT      01-AUG-1992 (Rel. 23, Created)
DT      01-AUG-1992 (Rel. 23, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      GENOME POLYPEPTIDE [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE      ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE      (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE      (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE      NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE      NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE      (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS      Hepatitis C virus (isolate Japanese) (HCV).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
CN      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE-91088550; PubMed-2175903;
RA      Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA      Sugimura T., Shimotohno K.;
RT      "Molecular cloning of the human hepatitis C virus genome from
RT      Japanese patients with non-A, non-B hepatitis.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RN      [2]
RN      DISCUSSION OF SEQUENCE.
RX      MEDLINE-91192160; PubMed-1849488;
RA      Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA      Ohkoshi S., Shimotohno K.;
RT      "Molecular structure of the Japanese hepatitis C viral genome.";
RL      FEBS Lett. 280:325-328(1991).
CC      -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC      HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC      NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA.
CC      -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in the way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; D90208; BAA14233.1; -
CC      PIR; A39253; GNWVCJ.
CC      HSP; P27958; LHET.
CC      MEROPS; S29.001; -
CC      MEROPS; U39.001; -
CC      INTERPRO; IPR000745; -
CC      INTERPRO; IPR001490; -
CC      INTERPRO; IPR002166; -
CC      INTERPRO; IPR002518; -
CC      INTERPRO; IPR002519; -
CC      INTERPRO; IPR002521; -
CC      INTERPRO; IPR002522; -
CC      INTERPRO; IPR002531; -

```

```

DR INTERPRO: IPR002868; .
DR PFAM: PF01560; HCV_NS1; 1.
DR DR PFAM: PF01538; HCV_NS2; 1.
DR DR PFAM: PF01006; HCV_NS4a; 1.
DR DR PFAM: PF01001; HCV_NS4b; 1.
DR DR PFAM: PF01506; HCV_NS5a; 1.
DR DR PFAM: PF00998; HCV_RdRP; 1.
DR DR PFAM: PF01543; HCV_Capsid; 1.
DR DR PFAM: PF01542; HCV_core; 1.
DR DR PFAM: PF01539; HCV_env; 1.
DR DR PFAM: PF01538; HCV_NS2; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural REMOVED FROM CAPSID PROTEIN C BY THE
INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT CHAIN 3011 369
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2788 2788
FT SEQUENCE 3010 AA; 327017 MW; AA993794F46DB185 CRC64;

Query Match 83.9%; Score 1253; DB 1; Length 3010;
Best Local Similarity 79.5%; Pred. No. 4.6e-100;
Matches 202; Conservative 25; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKONVQLINTNGSWHLNSTALNCNDLSLNTGLAGLYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GPSQKQLVNTNGSWHLNSTALNCNDLSLNTGLAGLYHHKFNSSGCPERLASCRPIDEF 465
QY 61 DQGWGPISTANGSGDQRPYCHYPPKPGCGIVPAKSVCGVCFPTSPVVGTTDRSGAP 120
DB 466 AQGWGPISTANGSGDQRPYCHYPPKPGCGIVPAKSVCGVCFPTSPVVGTTDRSGAP 525
QY 121 TYSWGENDTDFVFLNTRPPLGNFNGCTWMNSTGFTKVGCGAPCVTGAGNNTLHCPTDC 180
DB 526 TYSWGENDTDFVFLNTRPPLGNFNGCTWMNSTGFTKVGCGAPCVTGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSRCGSGPWITRCLVDYPIYLWHPYCTINTYIFKIRYVGVGVEHRLAACN 240
DB 586 FRKHDPATYSRCGSGPWITRCLVDYPIYLWHPYCTINTYIFKIRYVGVGVEHRLAACN 645
QY 241 WTRGERCDLEDRDR 254

```

Db 646 WTRGERCDLEDRDR 659

RESULT 6

POLG\_HCVJT STANDARD; PRT; 3010 AA.

```

AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE (EC 3.4.22.-); PROTEINASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RL infected individuals."
RL Virus Res. 23:39-53(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D11168; BAA01943.1; .
CC PIR: A45573; A45573.
CC HSSP: P27958; 1HEI.
CC MEROPS: S29.001; .
CC DR INTERPRO: IPR000745; .
CC DR INTERPRO: IPR001490; .
CC DR INTERPRO: IPR002166; .
CC DR INTERPRO: IPR002518; .
CC DR INTERPRO: IPR002519; .
CC DR INTERPRO: IPR002521; .
CC DR INTERPRO: IPR002522; .
CC DR INTERPRO: IPR002531; .
CC DR INTERPRO: IPR002868; .
CC DR PFAM: PF01560; HCV_NS1; 1.
CC DR PFAM: PF01538; HCV_NS2; 1.
CC DR PFAM: PF01006; HCV_NS4a; 1.
CC DR PFAM: PF01001; HCV_NS4b; 1.
CC DR PFAM: PF00998; HCV_RdRP; 1.
CC DR PFAM: PF01543; HCV_Capsid; 1.
CC DR PFAM: PF01542; HCV_core; 1.
CC DR PFAM: PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural REMOVED FROM CAPSID PROTEIN C BY THE
INIT_MET 1 1

```



```
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 384 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 729 NON-STRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1862 NON-STRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 NON-STRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2013 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT CHAIN 2014 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT CHAIN 347 TRANSMEM.
FT CHAIN 369 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1165 ATP (POTENTIAL).
FT SITE 1230 ATP (POTENTIAL).
FT SITE 1316 DECH BOX.
FT CARBOHYD 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Query Match 82.6%; Score 1234; DB 1; Length 3010;
Best Local Similarity 79.9%; Pred. No. 2e-98;
Matches 203; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

QY 1 GAKQVQLINTGSHLNTALNCDSLNTGLWLAGLYHHKFNSSCCPRLASCRPLTDF 60
DB 406 GPAQRLQINTGSHWINTALNCESLNTGFAALFYAHKFNSSCGPERMASCSIDKF 465
QY 61 DQGWGPIYANGSGDQRPYCYWHYPKPGCIYPAKSVGVPYCYFTSPVVGCTDRSGAP 120
DB 466 AQCWGPITVTEPRDLDRPYCYWHYAPQCGIYPAQSVQCVGVPYCYFTSPVVGCTDRSGAP 525
QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWNSGTFTKVCAGPCVIGGAGNNLHCPTDC 180
DB 526 TYNWGANETDVLNLRNTRPPQGNWFGCTWNSGTFTKTCGGPPCNGTGGVGNLTLCPTDC 585
QY 181 FRKHDPATYSRSGSGPWITPRLCVDPYPRLWHPCTINTYTIKIRMYGVGVHRELEAACN 240
DB 586 FRKHPEATYTKSGSGPWLTPRCIVDPYPRLWHPCTVNTFIPKVRMYGVGVHRELEAACN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 7
POLG_HCVJ8
AC POLG_HCVJ8 STANDARD; PRT; 3033 AA.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
```

```
DE DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
OS (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OC Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE-92230232; PubMed-1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: D10988; BAA01761.1; -
CC PIR: A40250; GNMVJ8.
CC HSSP: P27958; 1HEI.
CC
CC MEROPS: S29.001; -
CC INTERPRO: IPR000745; -
CC INTERPRO: IPR001490; -
CC INTERPRO: IPR002166; -
CC INTERPRO: IPR002518; -
CC INTERPRO: IPR002519; -
CC INTERPRO: IPR002521; -
CC INTERPRO: IPR002522; -
CC INTERPRO: IPR002531; -
CC INTERPRO: IPR002868; -
CC PFAM: PF01560; HCV_NS1; 1.
CC PFAM: PF01538; HCV_NS2; 1.
CC PFAM: PF01006; HCV_NS4a; 1.
CC PFAM: PF01001; HCV_NS4b; 1.
CC PFAM: PF01506; HCV_NS5a; 1.
CC PFAM: PF00998; HCV_RdRp; 1.
CC PFAM: PF01543; HCV_capsid; 1.
CC PFAM: PF01542; HCV_Core; 1.
CC PFAM: PF01539; HCV_env; 1.
CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding
CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
CC INIT_MET 1 1
CC REMOVED FROM CAPSID PROTEIN C BY THE
CC CELLULAR AMINOPEPTIDASE.
CC CAPSID PROTEIN C (POTENTIAL).
CC MATRIX PROTEIN (POTENTIAL).
CC MAJOR ENVELOPE PROTEIN E (POTENTIAL).
CC NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
CC NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
CC PROTEASE/HELICASE NS3 (POTENTIAL).
CC NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
CC NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
CC RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
CC TRANSMEM 347 369
CC ACT_SITE 1087 1087
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 733
CC CHAIN 734 1010
CC CHAIN 1011 1619
CC CHAIN 1620 1866
CC CHAIN 1867 2017
CC CHAIN 2018 3033
CC CHAIN 347 369
CC TRANSMEM 347 369
CC ACT_SITE 1087 1087
```

Tue Mar 6 12:02:35 2001

FT ACT\_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP\_BIND 1234 1241 ATP (POTENTIAL).

FT SITE 1320 1323 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 3033 3033 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 3033 AA; 330177 MW; 1A173E7E3381FD1A CRC64;

Query Match 78.3%; Score 1170; DB 1; Length 3033;

Best Local Similarity 76.7%; Pred. No. 6.6e-93;

Matches 198; Conservative 22; Mismatches 34; Indels 4; Gaps 2;

QY 1 GAKQNVQLINTNGSHLSTALNCNDSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60

DB 406 GAKQNVQLINTNGSHLSTALNCNDSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DOGWGPISYANG--SGDQRPYCHYPPKPGCGIYPAKSVGCVYCFPTSPVVVGTDRSG 118

DB 466 RIGWGTLEYETNTVNEEDMRPYCHYPPKPGCGIYPAKSVGCVYCFPTSPVVVGTDRSG 525

QY 119 APTYSWGENDTDFVLNNTTRPPLGNFNGCTWMNSTGFTKVCAGPCVIGGAGNNTLH--C 176

DB 586 PTDCFRKHPDATYSCGSGPWLTPCLVDYPRYLWHYPCNTVNTIFKRMVYGVGVEHRL 236

QY 237 AACNWTGRGCDLEDRDR 254

DB 646 AACNFTGRGCDLEDRDR 663

RESULT 8

POLG\_HCVJ7 STANDARD; PRT; 737 AA.

AC P27961;

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX

DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL

DE PROTEINS NS1 AND NS2] (FRAGMENT).

OS Hepatitis C virus (isolate HC-37) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=9220323; PubMed=1314459;

RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Lizuka H.,

RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.,

RT "Full-length sequence of a hepatitis C virus genome having poor

RT homology to reported isolates: comparative study of four distinct

RT genotypes.";

Virology 188:331-341(1992).

-!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

-!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; D10077; BAA00971.1; -

INTERPRO: IPR002519; -

INTERPRO: IPR002521; -

INTERPRO: IPR002522; -

INTERPRO: IPR002531; -

PFAM; PF01560; HCV\_NSI; 1.

PFAM; PF01543; HCV\_capsid; 1.

PFAM; PF01542; HCV\_core; 1.

PFAM; PF01539; HCV\_env; 1.

Polyprotein; Glycoprotein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein.

INIT\_MET 1 1

REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

CAPSID PROTEIN C (POTENTIAL).

MATRIX PROTEIN (POTENTIAL).

MAJOR ENVELOPE PROTEIN E (POTENTIAL).

NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

POTENTIAL.

TRANSMEM 347 369

CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).

NON\_TER 737 737

SQ SEQUENCE 737 AA; 81691 MW; 67DFAE11854122F2 CRC64;

Query Match 77.5%; Score 1158; DB 1; Length 737;

Best Local Similarity 75.6%; Pred. No. 1.6e-92;

Matches 195; Conservative 23; Mismatches 36; Indels 4; Gaps 2;

QY 1 GAKQNVQLINTNGSHLSTALNCNDSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60

DB 406 GSRQISLINTNGSHLSTALNCNDSLNTGLAGLFYHHKFNSSGCGPERLASCRPLTDF 465

QY 61 DOGWGPISYANG--SGDQRPYCHYPPKPGCGIYPAKSVGCVYCFPTSPVVVGTDRSG 118

DB 466 RIGWGTLEYETNTVNEEDMRPYCHYPPKPGCGIYPAKSVGCVYCFPTSPVVVGTDRSG 525

QY 119 APTYSWGENDTDFVLNNTTRPPLGNFNGCTWMNSTGFTKVCAGPCVIGGAGNNTLH--C 176

DB 526 VPTYWGENETDFVLNNTTRPPLGNFNGCTWMNSTGFTKVCAGPCVIGGAGNNTLH--C 585

QY 177 PTDCFRKHPDATYSCGSGPWLTPCLVDYPRYLWHYPCNTVNTIFKRMVYGVGVEHRL 236

Db 586 PTDCFRKHPDITYLKGAGPWLTPCLVDYPLRLWHYPCVTNFTIFKVMYVGGVEHRLD 645  
 QY 237 AACNWTGRGCDLEDRDR 254  
 Db 646 AACNFTGRGCDLEDRDR 663

RESULT 9  
 POLG\_HCVJ6 STANDARD; PRT; 3033 AA.  
 AC P26660;  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21);  
 DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);  
 DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);  
 DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)  
 DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate HC-J6) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92044440; PubMed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kural K., Lizuka H.,  
 RA Machida A., Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 RT from a human carrier: comparison with reported isolates for conserved  
 RL and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704(1991).  
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D00944; BAA00792.1; -  
 DR PIR: JQ1303; JQ1303.  
 DR HSSP: P27958; 1HE1.  
 DR MEROPS: S29.001; -  
 DR MEROPS: U39.001; -  
 DR INTERPRO: IPR000745; -  
 DR INTERPRO: IPR001490; -  
 DR INTERPRO: IPR002166; -  
 DR INTERPRO: IPR002518; -  
 DR INTERPRO: IPR002519; -  
 DR INTERPRO: IPR002521; -  
 DR INTERPRO: IPR002522; -  
 DR INTERPRO: IPR002531; -  
 DR INTERPRO: IPR002868; -  
 DR PFAM: PF01560; HCV\_NS1; 1.  
 DR PFAM: PF01538; HCV\_NS2; 1.  
 DR PFAM: PF01006; HCV\_NS4a; 1.  
 DR PFAM: PF01001; HCV\_NS4b; 1.  
 DR PFAM: PF01506; HCV\_NS5a; 1.  
 DR PFAM: PF00998; HCV\_RdRp; 1.  
 DR PFAM: PF01543; HCV\_Capsid; 1.  
 DR PFAM: PF01542; HCV\_core; 1.

DR KW PFAM: PF01539; HCV\_env; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural  
 FT INIT\_MET 1 1  
 FT CELLULAR AMINOPEPTIDASE.  
 FT CAPSID PROTEIN C (POTENTIAL).  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 733  
 FT CHAIN 734 1010  
 FT CHAIN 1011 1619  
 FT CHAIN 1620 1866  
 FT CHAIN 1867 2017  
 FT CHAIN 2018 3033  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1087 1087  
 FT ACT\_SITE 1111 1111  
 FT ACT\_SITE 1169 1169  
 FT NP\_BIND 1234 1241  
 FT SITE 1320 1323  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 477 477  
 FT CARBOHYD 534 534  
 FT CARBOHYD 542 542  
 FT CARBOHYD 558 558  
 FT CARBOHYD 578 578  
 FT CARBOHYD 627 627  
 FT CARBOHYD 649 649  
 FT CARBOHYD 1091 1091  
 FT CARBOHYD 2038 2038  
 FT CARBOHYD 2811 2811  
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 77.4%; Score 1156; DB 1; Length 3033;  
 Best Local Similarity 75.6%; Pred. No. 1.le-91;  
 Matches 195; Conservative 27; Mismatches 32; Indels 4; Gaps 3;  
 QY 1 GARQNVQLINTNGSHLSTALNCNDSLNTGLWGLAGLFYHHKFNSSGCCPERLASCRPLTDF 60  
 Db 406 GARQKIQLINTNGSHLSTALNCNDSLNTGLWGLAGLFYHHKFNSSGCCPERLASCRPLTDF 60  
 QY 61 DOGWGPISYA-NGSCP-DORPYCHWHYPPKPGIVPAKSVCGPVYCFPTSPVVVGTDRSG 118  
 Db 466 RVWGALQYEDNVTNPEDMRPYCHWHYPPRQCGVVSASSVCGPVYCFPTSPVVVGTDRIG 525  
 QY 119 APTYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGPVCVIGGAGNNT--LHC 176  
 Db 526 APTYTWGENETDVFLLNSTRPQGSWFGCTWMNSTGFTKVCAGPVCVIGGAGNNT--LHC 176  
 QY 177 PTDCFRKHPDATYSGCGSPWITPRCLVDYPLRLWHYPCVTNFTIFKVMYVGGVEHRLD 236  
 Db 586 PTDCFRKHPDITYLKGAGPWLTPCLVDYPLRLWHYPCVTNFTIFKVMYVGGVEHRLT 645  
 QY 237 AACNWTGRGCDLEDRDR 254  
 Db 646 AACNFTGRGCDLEDRDR 663  
 RESULT 10  
 POLG\_HCVJ5 STANDARD; PRT; 737 AA.  
 AC P27960;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)

QY	1	GA	KONVOLINTNGSWHLNSTALNCNDLSNTAGLFYHHKHNSSGCPBERLASCPRLTDF	60
Db	406	GS	KONQLINTNGSWHLNSTALNCNDLSNTGFIASLFYVNRNNSGCPHLRSLVCSIEAF	465
QY	61	DQ	WGPISTA-NGSGP-DORPWCYHPPKPCGIVPAKSVCGPVYCTPSPVVVGTTDRSG	118
Db	466	RI	CGWGLQVEDNVTNEDMRPYCWHYPPKPCGIVPARSVCGPVYCTPSPVVVGTTDARG	525
QY	119	AP	YSGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGN--NTLHC	176
Db	526	VP	YTGNETDVELLNSTRPPRGSWFGCTWMNSTGFTKTCGAPPCRIRADNFASTDLLC	585
QY	177	PT	CFRKHDPATYSRCGSGPWITPCRLVDYPYRLWHYPCITNVTIEKIRMYVGGVEHRLC	236
Db	586	PT	CFRKHSDATYIKCGSGPWLTPCKMDVPYRLWHYPCITNVTNIEKIRMYVGGVEHRLT	645
QY	237	AA	GNWTRGRCDLDRDR	254
Db	646	AA	GNFTRGDCPNLEDRDR	663

POLG\_HICV114  
TD POLG

OS	Hepat
OC	virus
OC	Hepat

CC the R  
CC use  
CC modifi  
CC entiti  
CC or so

FT  
CHAI

U-1: Gen. Virus: 012472  
SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

EMBL: D10688; BAA01530.1; --			
INTERPRO: IPR002519; --			
INTERPRO: IPR002521; --			
INTERPRO: IPR002522; --			
INTERPRO: IPR002531; --			
PFAM: PF01560; HCV_NS1; 1.			
PFAM: PF01543; HCV_capsid; 1.			
PFAM: PF01542; HCV_core; 1.			
PFAM: PF01539; HCV_env; 1.			
Polyprotein; Glycoprotein; Coat protein; Envelope protein; protein.			
Transmembrane; Nonstructural			
INIT_MET 1			
CHAIN	1	115	
CHAIN	116	191	
CHAIN	192	383	
CHAIN	384	>520	
			CELLULAR AMINOPEPTIDASE.
			CAPSID PROTEIN C (POTENTIAL).
			MATRIX PROTEIN (POTENTIAL).
			MAJOR ENVELOPE PROTEIN E (POTENTIAL).
			NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT TRANSMEM 347 369 BY SIMILARITY.  
 FT CARBOHYD 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 418 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 424 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 431 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 449 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 520 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 520 AA; 56499 MW; AAL35246CF20D525 CRC64;

Query Match 34.2%; Score 511; DB 1; Length 520;  
 Best Local Similarity 73.7%; Pred. No. 6.2e-37;  
 Matches 84; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHWLNSTALNCNDSLTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
 Db 407 GPSOKIQLINTNGSHWLNSTALNCNDSLTGWLGLFYHHKFNSSGCPERLASCRPIDKF 466  
 QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPCGIVPAKSVCGPYVCTPSPVVGTT 114  
 Db 467 DOGWGPVYAEPSISEQRPYCHWYAPRCGTIPASEVCGPYVCTPSPVVGTT 520

RESULT 12  
 POLG\_HCVHK STANDARD; PRT; 520 AA.  
 AC Q01403;  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1)] (FRAGMENT).  
 OS Hepatitis C virus (isolate HCV-RF) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93019030; PubMed=1383400;  
 RA Abe K., Inchausti G., Fujisawa K.;  
 RT "Genomic characterization and mutation rate of hepatitis C virus  
 RT isolated from a patient who contracted hepatitis during an epidemic  
 RT of non-A, non-B hepatitis in Japan."  
 RL J. Gen. Virol. 73:2725-2729(1992).  
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10687; BAA01529.1;  
 CC PIR: J01925; J01925;  
 CC INTERPRO: IPR002519;  
 CC INTERPRO: IPR002521;  
 CC INTERPRO: IPR002522;  
 CC INTERPRO: IPR002531;  
 CC PFAM: PF01560; HCV\_NSI.1;  
 CC PFAM: PF01543; HCV\_capsid.1;  
 CC PFAM: PF01542; HCV\_core.1;  
 CC PFAM: PF01339; HCV\_env.1;  
 KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;  
 KW Transmembrane; Nonstructural protein.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 >520  
 FT TRANSMEM 347 369  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 233 233  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 418 418  
 FT CARBOHYD 424 424  
 FT CARBOHYD 431 431  
 FT CARBOHYD 449 449  
 FT NON\_TER 520 520  
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FF27349B CRC64;

Query Match 34.2%; Score 511; DB 1; Length 520;  
 Best Local Similarity 73.7%; Pred. No. 6.2e-37;  
 Matches 84; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSHWLNSTALNCNDSLTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
 Db 407 GPSOKIQLINTNGSHWLNSTALNCNDSLTGWLGLFYHHKFNSSGCPERLASCRPIDKF 466  
 QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPCGIVPAKSVCGPYVCTPSPVVGTT 114  
 Db 467 DOGWGPVYAEPSISEQRPYCHWYAPRCGTIPASEVCGPYVCTPSPVVGTT 520

RESULT 13  
 POLG\_HCVJ2 STANDARD; PRT; 513 AA.  
 AC P27959;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);  
 DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2  
 DE (GP68) (GP70) (NS1)] (FRAGMENT).  
 OS Hepatitis C virus (isolate HC-J2) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 CC Hepacivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92230232; PubMed=1314459;  
 RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Mizuka H.,  
 RA Tanaka T., Fukuda S., Tada F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 RT homology to reported isolates: comparative study of four distinct  
 RT genotypes."  
 RL Virology 188:331-341(1992).  
 CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND MRNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D10074; BAA00968.1;  
 CC INTERPRO: IPR002519;  
 CC INTERPRO: IPR002521;

```
DR INTERPRO: IPR002522; -
DR INTERPRO: IPR002531; -
DR PFAM: PF01560; HCV_NSI; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 >513
FT TRANSMEM 347 369
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT NON_TER 513
SQ SEQUENCE 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;

Query Match 31.8%; Score 475; DB 1; Length 513;
Best Local Similarity 73.1%; Pred. No. 7.7e-34;
Matches 79; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 1 GAKQNVOLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKNSGCPERLASCRPLDIF 60
DB 406 GASQKQLINTNGSWHLNSTALNCNDSLNTGFLAALFYTHKFNASGCPERLASCRSDIGF 465

QY 61 DQGWGPISVANGSGPDRPCYHHPKPGIYPAKSVCGPYVCFTFSP 108
DB 466 DQGWGPITTEPGDSQKPCYHWHYAPQRCVSVSAADVCGPYVCFTFSP 513

RESULT 14
POLG_HCVTH STANDARD; PRT; 321 AA.
AC P27957;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: MATRIX PROTEIN (ENVELOPE PROTEIN M);
DE MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN NS1] (FRAGMENT).
OS Hepatitis C virus (isolate TH) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53136; CAA37296.1; -
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53134; CAA37294.1; -
DR INTERPRO: IPR002519; -
DR INTERPRO: IPR002521; -
DR INTERPRO: IPR002531; -
DR PFAM: PF01560; HCV_NSI; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01539; HCV_env; 1.
KW Polyprotein; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT NON_TER 1
FT CHAIN <1 75
FT CHAIN 76 267
FT CHAIN 268 >321
FT CARBOHYD 80 80
FT CARBOHYD 93 93
FT CARBOHYD 118 118
FT CARBOHYD 189 189
FT CARBOHYD 301 301
FT CARBOHYD 307 307
FT CARBOHYD 314 314
FT CARBOHYD 321 321
FT NON_TER 321
SQ SEQUENCE 321 AA; 34074 MW; B2E83F521C3B520 CRC64;

Query Match 11.6%; Score 173; DB 1; Length 321;
Best Local Similarity 90.6%; Pred. No. 5e-08;
Matches 29; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVOLINTNGSWHLNSTALNCNDSLNTGWL 32
DB 290 GARQNIQLINTNGSWHLNSTALNCNDSLNTGWL 321

RESULT 15
POLG_HCVEO STANDARD; PRT; 138 AA.
AC P27953;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35);
DE ENVELOPE GLYCOPROTEIN E2 (GP68) (GP70) (NS1)] (FRAGMENT).
OS Hepatitis C virus (isolate EC10) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91112009; PubMed=1846505;
RA Weiner A.J., Brauer M.J., Rosenblatt J., Richman K.H., Tung J.,
RA Crawford K., Bonino F., Saracco G., Choo Q.-L., Houghton M.,
RA Han J.H.;
RT "Variable and hypervariable domains are found in the regions of HCV
RT corresponding to the flavivirus envelope and NS1 proteins and the
RT pestivirus envelope glycoproteins."
RL Virology 180:842-848(1991).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53136; CAA37296.1; -
```

DR INTERPRO: IPR002519; ..  
 DR INTERPRO: IPR002531; ..  
 DR PFAM: PF01560; HCV\_NS1; 1.  
 DR PFAM: PF01539; HCV\_env; 1.  
 KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;  
 KW Transmembrane; Nonstructural protein.  
 FT NON\_TER 1  
 FT CHAIN <1 84  
 FT CHAIN 85 >138  
 FT CARBOHYD 6  
 FT CARBOHYD 118 118  
 FT CARBOHYD 124 124  
 FT CARBOHYD 131 131  
 FT NON\_TER 138 138  
 SQ SEQUENCE 138 AA: 14781 MW: 14781 MW: CD3FOA962DEABIAD CRC64;  
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 11.4%; Score 171; DB 1; Length 138;  
 Best Local Similarity 90.6%; Pred. No. 3e-08;  
 Matches 29; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAKONVOLINTNGSWHLNSTALNCNDSINTGW 32  
 DB 107 GAKONQIINTNGSWHLNSTALNCNDSINTGW 138

Search completed: March 6, 2001, 11:53:49  
 Job time: 186 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:55:05 ; Search time 75.33 Seconds  
(without alignments)  
395.205 Million cell updates/sec

Title: US-09-407-430-3  
Perfect score: 1494  
Sequence: 1 GAKQNVQLINTNGSWHLNST.....LEAACNWTGRCDELDLDR 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL15.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1494	100.0	3011	12 Q9IF5	Q9if5 hepatitis c
2	1438	96.3	3011	12 Q03463	Q03463 hepatitis c
3	1426	95.4	778	12 Q04184	Q04184 h genome po
4	1418	94.9	3011	12 Q36579	Q36579 hepatitis c
5	1418	94.9	3011	12 Q36608	Q36608 hepatitis c
6	1414	94.6	778	12 Q04185	Q04185 h genome po
7	1414	94.6	3011	12 Q36610	Q36610 hepatitis c
8	1393	93.2	3011	12 Q36609	Q36609 hepatitis c
9	1367	91.5	403	12 Q9PX22	Q9px22 hepatitis c
10	1304	87.3	3010	12 Q9J3H0	Q9j3h0 hepatitis c
11	1291	86.4	3011	12 Q81754	Q81754 hepatitis c
12	1285	86.0	3010	12 Q81760	Q81760 hepatitis c
13	1284	85.9	3010	12 Q9QIY6	Q9qiY6 hepatitis c
14	1284	85.9	3010	12 Q9QIY5	Q9qiY5 hepatitis c
15	1283	85.9	3010	12 Q9J3G4	Q9j3g4 hepatitis c
16	1282	85.8	3010	12 Q93016	Q93016 hepatitis c
17	1282	85.8	3010	12 Q9QIX5	Q9qiX5 hepatitis c
18	1281	85.7	3010	12 Q9QIX6	Q9qiX6 hepatitis c
19	1278	85.5	3010	12 Q9J3G1	Q9j3g1 hepatitis c

20	1277.5	85.5	3014	12	Q91936	hepatitis c
21	1276	85.4	3010	12	Q02828	h genome po
22	1276	85.4	3010	12	Q9J3G9	hepatitis c
23	1275	85.3	3010	12	Q81989	hepatitis c
24	1275	85.3	3010	12	Q9J3H9	hepatitis c
25	1275	85.3	3010	12	Q9J3H1	hepatitis c
26	1273	85.2	3010	12	Q901X4	hepatitis c
27	1272	85.1	640	12	Q68966	hepatitis c
28	1272	85.1	3010	12	Q90193	hepatitis c
29	1272	85.1	3010	12	Q90194	hepatitis c
30	1272	85.1	3010	12	Q9QIY9	hepatitis c
31	1272	85.1	3010	12	Q9J3H8	hepatitis c
32	1271	85.1	3010	12	Q9QIX3	hepatitis c
33	1271	85.1	3010	12	Q9J3G0	hepatitis c
34	1269	84.9	3010	12	Q02829	hepatitis c
35	1269	84.9	3010	12	Q9Q6P1	h genome po
36	1269	84.9	3010	12	Q9J3G8	hepatitis c
37	1268	84.9	3010	12	Q9Q120	hepatitis c
38	1268	84.9	3010	12	Q9J3I1	hepatitis c
39	1268	84.9	3010	12	Q9J3G2	hepatitis c
40	1267	84.8	3010	12	Q9J3H7	hepatitis c
41	1266	84.7	3010	12	Q68533	hepatitis c
42	1266	84.7	3010	12	Q9J3H3	hepatitis c
43	1266	84.7	3010	12	Q9J3G3	hepatitis c
44	1265.5	84.7	3015	12	Q9WPH5	hepatitis c
45	1265	84.7	3010	12	Q93077	hepatitis c

ALIGNMENTS

RESULT 1

Q9IF5  
ID Q9IF5 PRELIMINARY; PRT; 3011 AA.  
AC Q9IF5;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE POLYPROTEIN.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.;  
RT "An Infectious Clone of the HCV-1 Prototype Sequence."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF271632; AAF81759.1; -  
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;

Query Match 100.0%; Score 1494; DB 12; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 3.6e-139;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gap 0;

Qy	1	GAKQNVQLINTNGSWHLNSTALNCNDSLTGWLACLFYHHKFNSSGCCPERLASCRPLTDF	6
Db	406	GAKQNVQLINTNGSWHLNSTALNCNDSLTGWLACLFYHHKFNSSGCCPERLASCRPLTDF	45
Qy	61	DQGWGPISYANGSGDPORPCYCHHPKPCGIVPAKSVCGPYCYCTPSPVVVGTTRDSGAP	100
Db	466	DQGWGPISYANGSGDPORPCYCHHPKPCGIVPAKSVCGPYCYCTPSPVVVGTTRDSGAP	526
Qy	121	TYSWGENTDYFVLNTRPPLGNFECCTWMNSTGFTKVCAGPCPVIGAGNNTLHCPTDC	180
Db	526	TYSWGENTDYFVLNTRPPLGNFECCTWMNSTGFTKVCAGPCPVIGAGNNTLHCPTDC	58
Qy	181	FRKHPDATYSRCGGSPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRELAACN	24
Db	586	FRKHPDATYSRCGGSPWITPRCLVDYPRYLWHYPCTINYTIKIRMYVGGVEHRELAACN	64
Qy	241	WTRGERCDLEDRDR	254



```

QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 180
Db 526 TYSWGANDDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYIFKIRMYVGGVGEHRLAECN 240
Db 586 FRKHDPATYSRCGSGPWITPRCWDYPYRLWHYPCTINTYIFKIRMYVGGVGEHRLAECN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659

RESULT 4
Q36579 ID O36579 PRELIMINARY: PRT: 3011 AA.
AC O36579;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA.";
RL Science 277:570-574(1997).
DR EMBL: AF009606; AAB66324.1; -
DR INTERPRO: IPR000745; -
DR INTERPRO: IPR001410; -
DR INTERPRO: IPR001490; -
DR INTERPRO: IPR002166; -
DR INTERPRO: IPR002518; -
DR INTERPRO: IPR002519; -
DR INTERPRO: IPR002521; -
DR INTERPRO: IPR002522; -
DR INTERPRO: IPR002531; -
DR INTERPRO: IPR002868; -
DR PFAM: PF00998; HCV_RdRp; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01560; HCV_NS1; 1.
DR PRODOM: PD186062; -; 1.
KW Polyprotein.
SQ SEQUENCE 3011 AA; 327184 MW; E2E0E809C63C1B9 CRC64;

```

```

Query Match 94.9%; Score 1418; DB 12; Length 3011;
Best Local Similarity 94.1%; Pred. NO. 1.2e-131;
Matches 239; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 60
Db 406 GAKQNLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 465
QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 525

```

```

Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 180
Db 526 TYSWGANDDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGVIGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYIFKIRMYVGGVGEHRLAECN 240
Db 586 FRKHDPATYSRCGSGPWITPRCWDYPYRLWHYPCTINTYIFKIRMYVGGVGEHRLAECN 645
QY 241 WTRGERCDLEDRDR 254
Db 646 WTRGERCDLEDRDR 659

RESULT 5
Q36608 ID O36608 PRELIMINARY: PRT: 3011 AA.
AC O36608;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97385173; PubMed=9238047;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RA "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
DR EMBL: AF011751; AAB67036.1; -
DR INTERPRO: IPR000745; -
DR INTERPRO: IPR001410; -
DR INTERPRO: IPR001490; -
DR INTERPRO: IPR002166; -
DR INTERPRO: IPR002518; -
DR INTERPRO: IPR002519; -
DR INTERPRO: IPR002521; -
DR INTERPRO: IPR002522; -
DR INTERPRO: IPR002531; -
DR INTERPRO: IPR002868; -
DR PFAM: PF00998; HCV_RdRp; 1.
DR PFAM: PF01001; HCV_NS4b; 1.
DR PFAM: PF01006; HCV_NS4a; 1.
DR PFAM: PF01506; HCV_NS5a; 1.
DR PFAM: PF01538; HCV_NS2; 1.
DR PFAM: PF01539; HCV_env; 1.
DR PFAM: PF01542; HCV_core; 1.
DR PFAM: PF01543; HCV_capsid; 1.
DR PFAM: PF01560; HCV_NS1; 1.
DR PRODOM: PD186062; -; 1.
KW Polyprotein.
SQ SEQUENCE 3011 AA; 327114 MW; 0B75E6B81CB5C198 CRC64;

```

```

Query Match 94.9%; Score 1418; DB 12; Length 3011;
Best Local Similarity 94.1%; Pred. NO. 1.2e-131;
Matches 239; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 60
Db 406 GAKQNLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCCPERLASCRPLTDF 465
QY 61 DOGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120
Db 466 AOGWGPISYANGSGGLDERPYCHWYPPRPGIIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 525

```

Best Local Similarity 92.1%; Pred. No. 7.1e-132;		Matches 234; Conservative 11; Mismatches 9; Indels 0; Gaps 0;	
QY	121	TYSGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC	180
DB	526	TYSGANDTDVFLNNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC	585
QY	181	FKHPDATYSRCGSPWITPCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHLEAACN	240
DB	586	FRKHEATYSRCGSPWITPCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHLEAACN	645
QY	241	WTRGERCDLEDRDR	254
DB	646	WTRGERCDLEDRDR	659
RESULT 6			
ID	Q04185	PRELIMINARY;	PRT; 778 AA.
AC	Q04185; Q81811;		
DT	01-NOV-1998 (TREMBLrel. 08, Created)		
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C		
DE	(CORE PROTEIN); MATRIX PROTEIN		
DE	(ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEIN		
DE	NS1; NONSTRUCTURAL PROTEIN NS2; NONSTRUCTURAL PROTEIN NS4A;		
DE	NONSTRUCTURAL PROTEIN NS4B; HELICASE		
DE	(NS3); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (NS5)] (FRAGMENT).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91195357; PubMed=1849654;		
RA	Ogata N., Alter H.J., Miller R.H., Purcell R.H.;		
RT	"Nucleotide sequence and mutation rate of the H strain of hepatitis C		
RT	virus."		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:3392-3396(1991).		
CC	-1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE		
CC	HYPOTHETIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3		
CC	AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.		
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A		
CC	LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:		
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF		
CC	PROTEIN C AND RNA.		
CC	EMBL; M62382; AAB02128.1; -		
DR	INTERPRO: IPR002518; -		
DR	INTERPRO: IPR002519; -		
DR	INTERPRO: IPR002521; -		
DR	INTERPRO: IPR002522; -		
DR	INTERPRO: IPR002531; -		
DR	PFAM; PF01538; HCV_NS2; 1.		
DR	PFAM; PF01539; HCV_env; 1.		
DR	PFAM; PF01542; HCV_core; 1.		
DR	PFAM; PF01543; HCV_capsid; 1.		
DR	PFAM; PF01560; HCV_NS1; 1.		
KW	Polyprotein; Glycoprotein; RNA-directed RNA polymerase; Core protein;		
KW	Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane;		
KW	Nonstructural protein.		
FT	INIT_MET 1	REMOVED FROM CAPSID PROTEIN C BY THE	
FT	CHAIN 2	CELLULAR AMINOPEPTIDASE.	
FT	CHAIN 116	CAPSID PROTEIN C (POTENTIAL).	
FT	CHAIN 191	MATRIX PROTEIN (POTENTIAL).	
FT	CHAIN 192	MAJOR ENVELOPE PROTEIN E (POTENTIAL).	
FT	CHAIN 384	NONSTRUCTURAL PROTEIN NS1/E2	
FT	CHAIN 729	(POTENTIAL).	
FT	CHAIN 730	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).	
FT	CHAIN 347	POTENTIAL.	
FT	CHAIN 369		
FT	NON_TER 778		
FT	SEQUENCE 778 AA; 85530 MW; 6E37E50AB97B6C32 CRC64;		
Query Match 94.6%; Score 1414; DB 12; Length 778;			
Best Local Similarity 93.7%; Pred. No. 3e-131;			
Best Local Similarity 92.1%; Pred. No. 7.1e-132;		Matches 234; Conservative 11; Mismatches 9; Indels 0; Gaps 0;	
QY	1	GAKQVQLINTNSWHLNTALNCNDSLNTGWLGLFHHKFNSSGCPERLASCRPLTDF	60
DB	406	GPKQNIQLIKTNSWHLNTALNCNDSLNTGWLGLFHHKFNSSGCPERLASCRPLTDF	465
QY	61	DOGWGPISYANGSGDPORPCYCHYPPKPCGIVPAKSVCGVPVYCFTPSPVVVGTTDRSGAP	120
DB	466	DOGWGPISYANGSGDPORPCYCHYPPKPCGIVPAKSVCGVPVYCFTPSPVVVGTTDRSGAP	525
QY	121	TYSGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC	180
DB	526	TYNGENDTDVLLNNTRPPLGNWFGCTWMNSTGFTKVCAPPVCGAGNNTLHCPTDC	585
QY	181	FRKHPDATYSRCGSPWITPCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHLEAACN	240
DB	586	FRKHEATYSRCGSPWITPCLVDYPRYLWHYPCTINVTIFKIRMYVGVGVEHLEAACN	645
QY	241	WTRGERCDLEDRDR	254
DB	646	WTRGERCDLEDRDR	659
RESULT 7			
ID	O36610	PRELIMINARY;	PRT; 3011 AA.
AC	O36610;		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	POLYPROTEIN.		
OS	Hepatitis C virus strain H77.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
OX	NCBI_TaxID=63746;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H77;		
RX	MEDLINE=97385173; PubMed=9238047;		
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;		
RT	"Transcripts from a single full-length cDNA clone of hepatitis C virus		
RT	are infectious when directly transfected into the liver of a		
RT	chimpanzee."		
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).		
DR	EMBL; AF011753; AAB67038.1; -		
DR	HSSP; P27958; 1HEI.		
DR	INTERPRO: IPR000745; -		
DR	INTERPRO: IPR001410; -		
DR	INTERPRO: IPR001490; -		
DR	INTERPRO: IPR002166; -		
DR	INTERPRO: IPR002518; -		
DR	INTERPRO: IPR002519; -		
DR	INTERPRO: IPR002521; -		
DR	INTERPRO: IPR002522; -		
DR	INTERPRO: IPR002531; -		
DR	INTERPRO: IPR002868; -		
DR	PFAM; PF00998; HCV_RdRP; 1.		
DR	PFAM; PF01001; HCV_NS4b; 1.		
DR	PFAM; PF01006; HCV_NS4a; 1.		
DR	PFAM; PF01506; HCV_NS5a; 1.		
DR	PFAM; PF01538; HCV_NS2; 1.		
DR	PFAM; PF01539; HCV_env; 1.		
DR	PFAM; PF01542; HCV_core; 1.		
DR	PFAM; PF01543; HCV_capsid; 1.		
DR	PFAM; PF01560; HCV_NS1; 1.		
DR	PRODOM; PD186062; -; 1.		
KW	Polyprotein.		
QY	SEQUENCE 3011 AA; 327222 MW; 293F91430A0D4067 CRC64;		
Query Match 94.6%; Score 1414; DB 12; Length 3011;			

Matches 238; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 Db 406 GAKQNTQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCPERLASCRRLTDF 465

QY 61 DOGWGPISYANGSGDQRPYCHWHYPPKPGCGIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120  
 Db 466 AOGWGPISYANGSGLDPERPYCHWHYPPKPGCGIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 525

QY 121 TYSWGENDTDVFLNNTTRPPLGNWFGCTWMNSTGTFTKVGCGAPPCVIGGAGNNTLHCPTDC 180  
 Db 526 TYSWGANDTDVFLNNTTRPPLGNWFGCTWMNSTGTFTKVGCGAPPCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCNTINYTFKIRMYVGGVEHRLAECN 240  
 Db 586 FRKHPEATYSRCGSGPWITPRCMVDYPRYLWHYPCNTINYTFKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254  
 Db 646 WTRGERCDLEDRDR 659

RESULT 8

Q36609 ID Q36609 PRELIMINARY; PRT; 3011 AA.

AC Q36609;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DE POLYPROTEIN.

DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

OS Hepatitis C virus strain H77.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=63746;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-H77;

RX MEDLINE-97385173; PubMed-9238047;

RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;

RT "Transcripts from a single full-length cDNA clone of hepatitis C virus

RT are infectious when directly transfectected into the liver of a

RT chimpanzee."

RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).

DR EMBL: AF011752; AAB67037.1; -;

DR HSP: P27958; JHEI.

DR INTERPRO: IPR000745; -;

DR INTERPRO: IPR001410; -;

DR INTERPRO: IPR001490; -;

DR INTERPRO: IPR002166; -;

DR INTERPRO: IPR002518; -;

DR INTERPRO: IPR002519; -;

DR INTERPRO: IPR002521; -;

DR INTERPRO: IPR002522; -;

DR INTERPRO: IPR002531; -;

DR INTERPRO: IPR002868; -;

DR PFAM: PF00998; HCV\_RdRP; 1.

DR PFAM: PF01001; HCV\_NS4B; 1.

DR PFAM: PF01006; HCV\_NS4A; 1.

DR PFAM: PF01506; HCV\_NS5A; 1.

DR PFAM: PF01538; HCV\_NS2; 1.

DR PFAM: PF01539; HCV\_env; 1.

DR PFAM: PF01542; HCV\_core; 1.

DR PFAM: PF01543; HCV\_capsid; 1.

DR PRODOM: PD186062; -; 1.

KW Polyprotein.

SQ SEQUENCE 3011 AA; 327262 MW; 10DIC9702CA9B5DC CRC64;

Query Match 93.2%; Score 1393; DB 12; Length 3011;  
 Best Local Similarity 92.9%; Pred. No. 3.6e-129;  
 Matches 236; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 Db 406 GAKQNTQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCPERLASCRRLTDF 465

QY 61 DOGWGPISYANGSGDQRPYCHWHYPPKPGCGIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120  
 Db 466 AOGWGPISYANGSGLDPERPYCHWHYPPKPGCGIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 525

QY 121 TYSWGENDTDVFLNNTTRPPLGNWFGCTWMNSTGTFTKVGCGAPPCVIGGAGNNTLHCPTDC 180  
 Db 526 TYSWGANDTDVFLNNTTRPPLGNWFGCTWMNSTGTFTKVGCGAPPCVIGGAGNNTLHCPTDC 585

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCNTINYTFKIRMYVGGVEHRLAECN 240  
 Db 586 FRKHPEATYSRCGSGPWITPRCMVDYPRYLWHYPCNTINYTFKIRMYVGGVEHRLAECN 645

QY 241 WTRGERCDLEDRDR 254  
 Db 646 WTRGERCDLEDRDR 659

RESULT 9

Q9PX22 ID Q9PX22 PRELIMINARY; PRT; 403 AA.

AC Q9PX22;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)

DE E2/NS1 REGION.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-92348860; PubMed-1668329;

RA Kramsdoerf D., Porchon C., Brechot C.;

RT "Hepatitis C virus (HCV)-RNA in non-A, non-B chronic hepatitis in

RL J. Hepatol. 13:24-24(1991).

DR INTERPRO: IPR002519; -;

DR INTERPRO: IPR002531; -;

DR PFAM: PF01539; HCV\_env; 1.

DR PFAM: PF01560; HCV\_NS1; 1.

SQ SEQUENCE 403 AA; 44565 MW; 06278192EAC5B3F1 CRC64;

Query Match 91.5%; Score 1367; DB 12; Length 403;  
 Best Local Similarity 87.4%; Pred. No. 1.6e-127;  
 Matches 222; Conservative 25; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 Db 73 GAKQNTQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCPERLASCRPLAD 132

QY 61 DOGWGPISYANGSGDQRPYCHWHYPPKPGCGIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120  
 Db 133 DOGWGPISYANGSGDQRPYCHWHYPPKPGCGIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 192

QY 121 TYSWGENDTDVFLNNTTRPPLGNWFGCTWMNSTGTFTKVGCGAPPCVIGGAGNNTLHCPTDC 180  
 Db 193 TYSWGANDTDVFLNNTTRPPLGNWFGCTWMNSTGTFTKVGCGAPPCVIGGAGNNTLHCPTDC 252

QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCNTINYTFKIRMYVGGVEHRLAECN 240  
 Db 253 FRKHPEATYSRCGSGPWITPRCLVDYPRYLWHYPCNTINYTFKIRMYVGGVEHRLAECN 312

QY 241 WTRGERCDLEDRDR 254  
 Db 313 WTRGERCDLEDRDR 326

```

RESULT 10
Q9J3H0 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD22;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF207763; AAF65953.1;
SQ SEQUENCE 3010 AA; 327098 MW; 737EEF31E3C2B28D CRC64;

Query Match 87.3%; Score 1304; DB 12; Length 3010;
Best Local Similarity 83.9%; Pred. No. 2.4e-120; Indels 0; Gaps
Matches 213; Conservative 19; Mismatches 22;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
QY 61 DQGWGPISYANGSGDPQRYCWHYPPKCGIVPAKSCVCGVYCFPTSPVVGTTDRSGAP 120
DB 466 DQGWGPIITTEPNSPQRYCWHYAPRPGCIVPASGCVGVPYCFPTSPVVGTTDRSGVP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVITGGAGNNTLHCPTDC 180
DB 526 TYSWGENETDVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVITGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWITPCLVDYPRYLWHYPCNTINTYFKIRMYVGVGVEHRLAECN 240
DB 586 FRKHDPATYSCGSGPWITPCLVDYPRYLWHYPCNTINTYFKIRMYVGVGVEHRLAECN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 11
Q81754 PRELIMINARY; PRT; 3011 AA.
AC Q81754;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RA Okamoto H.;
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HC-G9;
RX MEDLINE=94172337; PubMed=8126459;
RA Okamoto H., Kojima M., Sakamoto M., Iizuka H., Hadiwandowo S.,
RA Suwignyo S., Miyakawa Y., Mayumi M.;
RT "The entire nucleotide sequence and classification of a hepatitis C
virus isolate of a novel genotype from an Indonesian patient with
```

```

chronic liver disease.";
J. Gen. Virol. 75:629-635(1994).
DR EMBL: D14853; BAA03581.1;
DR HSSE; P26663; LJXP.
DR MEROPS; S29.001;
DR INTERPRO; IPR000745;
DR INTERPRO; IPR001410;
DR INTERPRO; IPR001490;
DR INTERPRO; IPR002186;
DR INTERPRO; IPR002518;
DR INTERPRO; IPR002519;
DR INTERPRO; IPR002521;
DR INTERPRO; IPR002522;
DR INTERPRO; IPR002531;
DR INTERPRO; IPR002868;
DR PFAM; PF00998; HCV_RDRP; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01506; HCV_NS5a; 1.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01542; HCV_core; 1.
DR PFAM; PF01543; HCV_capsid; 1.
DR PFAM; PF01560; HCV_NS1; 1.
DR PRODOM; PD186062; -; 1.
KW Polyprotein.
ET CHAIN 1 191 CORE.
ET CHAIN 192 383 E1.
FT CHAIN 384 809 E2/NS1.
FT CHAIN 810 1006 NS2.
FT CHAIN 1007 1657 NS3.
FT CHAIN 1658 1972 NS4.
FT CHAIN 1973 3011 NS5.
SQ SEQUENCE 3011 AA; 327213 MW; 9C16C120F4E79268 CRC64;

Query Match 86.4%; Score 1291; DB 12; Length 3011;
Best Local Similarity 83.1%; Pred. No. 4.6e-119; Indels 0; Gaps
Matches 211; Conservative 23; Mismatches 20;

QY 1 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 406 GAKQNVQLINTNGSHLNTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60
QY 61 DQGWGPISYANGSGDPQRYCWHYPPKCGIVPAKSCVCGVYCFPTSPVVGTTDRSGAP 120
DB 466 DQGWGPIITTEPNSPQRYCWHYAPRPGCIVPASGCVGVPYCFPTSPVVGTTDRSGVP 525
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVITGGAGNNTLHCPTDC 180
DB 526 TYRWGANETDVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVITGGAGNNTLHCPTDC 585
QY 181 FRKHDPATYSCGSGPWITPCLVDYPRYLWHYPCNTINTYFKIRMYVGVGVEHRLAECN 240
DB 586 FRKHDPATYSCGSGPWITPCLVDYPRYLWHYPCNTINTYFKIRMYVGVGVEHRLAECN 645
QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCDLEDRDR 659

RESULT 12
Q81760 PRELIMINARY; PRT; 3010 AA.
ID Q81760
AC Q81760;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN PRECURSOR.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
```

OX NCBI\_TaxID=11103;  
 [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=HC-C2;

RA MEDLINE=93359897; PubMed=8394876;  
 RA Wang Y., Okamoto H., Tsuda F., Nagayama R., Tao Q.M., Mishiro S.;

RT "Prevalence, genotypes, and an isolate (HC-C2) of hepatitis C virus in  
 Chinese patients with liver disease.";  
 RL J. Med. Virol. 40:254-260(1993).

DR EMBL: D10934; BAA01728.1; -;  
 DR HSSP: P26663; JNS3

DR INTERPRO: IPR000745; -;  
 DR INTERPRO: IPR001410; -;

DR INTERPRO: IPR001490; -;  
 DR INTERPRO: IPR002166; -;

DR INTERPRO: IPR002518; -;  
 DR INTERPRO: IPR002519; -;

DR INTERPRO: IPR002521; -;  
 DR INTERPRO: IPR002522; -;

DR INTERPRO: IPR002868; -;  
 DR PFAM: PF00998; HCV\_RdRP; 1.

DR PFAM: PF01001; HCV\_NS4b; 1.  
 DR PFAM: PF01006; HCV\_NS4a; 1.

DR PFAM: PF01506; HCV\_NS3a; 1.  
 DR PFAM: PF01538; HCV\_NS2; 1.

DR PFAM: PF01539; HCV\_env; 1.  
 DR PFAM: PF01542; HCV\_core; 1.

DR PFAM: PF01543; HCV\_capsid; 1.  
 DR PFAM: PF01560; HCV\_NS1; 1.

DR PRODOM: PD186062; -; 1.  
 KW Polyprotein.

SQ SEQUENCE 3010 AA; 326857 MW; EA7D306A4BA2E224 CRC64;

Query Match

Best Local Similarity 86.0%; Score 1285; DB 12; Length 3010;  
 Matches 208; Conservative 21; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 DB 406 GPSQKIQLVNTNGSWHLNSTALNCNDSFNTGLAALFYAHRFNSSGCPERMASCRSIDKF 465

QY 61 DQGWGPISYANGSGDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120  
 DB 466 DQGWGPIYQGDSPDQRPYCHYHPKPCGIVPASEVCGPVYCFPTSPVVGTTDRSGAP 525

QY 121 TYSWGENDTDFVLLNTRPPLGNWFCGCTWMNSTGFTKVCAGAPCVIGGAGNNTLHCPTDC 180  
 DB 526 TYNWGENETDVLNTRPQGNWFCGCTWMNSTGFTKCGGPPCNGAGNNTLHCPTDC 585

QY 181 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYFIKIRMYGGVHRLEAACN 240  
 DB 586 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYFIKIRMYGGVHRLEAACN 645

QY 241 WTRGERCDLEDRDR 254  
 DB 646 WTRGERCDLEDRDR 659

RESULT 13

QYQIY6

ID QYQIY6 PRELIMINARY; PRT: 3010 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=11103;  
 RN [1]

RP

RC SEQUENCE FROM N.A.  
 RX STRAIN=MD3-1;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,  
 RA Tazawa J.I., Izumi N., Marumo F., Sato C.;

RT "Time-related changes in full-length hepatitis C virus and hepatitis  
 activity.";  
 RL Virology 263:244-253(1999).

RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=MD3-1;  
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,

RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF165049; AAD56184.1; -;  
 DR HSSP: P26663; LJXP.

DR INTERPRO: IPR000745; -;  
 DR INTERPRO: IPR001410; -;

DR INTERPRO: IPR002166; -;  
 DR INTERPRO: IPR002518; -;

DR INTERPRO: IPR002531; -;  
 DR PFAM: PF00998; HCV\_RdRP; 1.

DR PFAM: PF01006; HCV\_NS4a; 1.  
 DR PFAM: PF01538; HCV\_NS2; 1.

DR PFAM: PF01539; HCV\_env; 1.  
 DR PFAM: PF01560; HCV\_NS1; 1.

KW Polyprotein.

SQ SEQUENCE 3010 AA; 327369 MW; 998C7F293EAAEC8D CRC64;

Query Match

Best Local Similarity 85.9%; Score 1284; DB 12; Length 3010;  
 Matches 210; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGLAGLFYHHKFNSSGCPERLASCRPLTDF 60  
 DB 406 GASQKIQLVNTNGSWHLNSTALNCNDSLNTGLAALFYTHKFNSSGCPERMASCRSIDKF 465

QY 61 DQGWGPISYANGSGDQRPYCHYHPKPCGIVPAKSVCGPVYCFPTSPVVGTTDRSGAP 120  
 DB 466 DQGWGPIYQGDSPDQRPYCHYHPKPCGIVPASEVCGPVYCFPTSPVVGTTDRSGAP 525

QY 121 TYSWGENDTDFVLLNTRPPLGNWFCGCTWMNSTGFTKVCAGAPCVIGGAGNNTLHCPTDC 180  
 DB 526 TYSWGENETDVLNTRPQGNWFCGCTWMNSTGFTKCGGPPCNGAGNNTLHCPTDC 585

QY 181 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYFIKIRMYGGVHRLEAACN 240  
 DB 586 FRKHDPATYRCGSGPWITPRCLVDYPYRLWHYPCNTINTYFIKIRMYGGVHRLEAACN 645

QY 241 WTRGERCDLEDRDR 254  
 DB 646 WTRGERCDLEDRDR 659

RESULT 14

QYQIY5

ID QYQIY5 PRELIMINARY; PRT: 3010 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE POLYPROTEIN.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=11103;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD3-2;

RX MEDLINE=20013325; PubMed=10544098;

RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,

```
RA Tazawa J.i., Izumi N., Marumo F., Sato C.;
RT "Time-related changes in full-length hepatitis C virus and hepatitis
RL activity.";
RL Virology 263:244-253(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MD3-2;
RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF165050; AAD56185.1;
DR HSP; P28663; IJXP.
DR INTERPRO; IPR000745;
DR INTERPRO; IPR001410;
DR INTERPRO; IPR002166;
DR INTERPRO; IPR002518;
DR INTERPRO; IPR002519;
DR INTERPRO; IPR002531;
DR PFAM; PF00998; HCV_RdRP; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
DR PFAM; PF01538; HCV_NS2; 1.
DR PFAM; PF01539; HCV_env; 1.
DR PFAM; PF01560; HCV_NS1; 1.
KW Polyprotein
SQ SEQUENCE 3010 AA; 327431 MW; 15190E3463DEB8C35 CRC64;

Query Match 85.9%; Score 1284; DB 12; Length 3010;
Best Local Similarity 82.7%; Pred. No. 2.3e-118;
Matches 210; Conservative 17; Mismatches 27; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
DB 406 GASQRIQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRSIDTF 465

QY 61 DOGWGPIYANGSGDPQRPYCHWHYPPKCGIVPAKSVCGPVYCFTPSPVVVGTTRDSGAP 120
DB 466 DOGWGPIYANGSGDPQRPYCHWHYAPPCGVVPAVEVCGPVYCFTPSPVVVGTTRDFGVP 525

QY 121 TVSWGENDTVFVLNTRPPLGNWFGCTWMNSTGFTKVGAPPCVIGGAGNNTLHCPTDC 180
DB 526 TYNWGENETDVLILNTRPPQGNWFGCTWMNGTGFTKTCGGPPCNIGGAGNNTLTCPTDC 585

QY 181 FRKHDPATYSRCGSPWITPRCLVDYPYRLWHYPCTINVTIFKIRMYVGVGVEHRLAAACN 240
DB 586 FRKHPEATYTKCGSPWLTPRCLVDYPYRLWHYPCTVNETTKVRYMYGVGVEHRLVAAACN 645

QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCNLEDRDR 659

RESULT 15
Q9J3G4 PRELIMINARY; PRT; 3010 AA.
AC Q9J3G4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MD28;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF207769; AAF65959.1;
SQ SEQUENCE 3010 AA; 327318 MW; 0595728AC62464F3 CRC64;
```

```
Query Match 85.9%; Score 1283; DB 12; Length 3010;
Best Local Similarity 82.3%; Pred. No. 2.8e-118;
Matches 209; Conservative 20; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRPLTDF 60
DB 406 GASQRIQLINTNGSWHLNSTALNCNDSLNTGWLGLAGLFYHHKFNSSGCGPERLASCRSIDTF 465

QY 61 DOGWGPIYANGSGDPQRPYCHWHYPPKCGIVPAKSVCGPVYCFTPSPVVVGTTRDSGAP 120
DB 466 DOGWGPIYAKSGSPQRPYCHWHYAPPCGVVPAVEVCGPVYCFTPSPVVVGTTRDFGAP 525

QY 121 TVSWGENDTVFVLNTRPPLGNWFGCTWMNSTGFTKVGAPPCVIGGAGNNTLHCPTDC 180
DB 526 TYNWGENETDVLILNTRPPQGNWFGCTWMNGTGFTKTCGGPPCNIGGAGNNTLTCPTDC 585

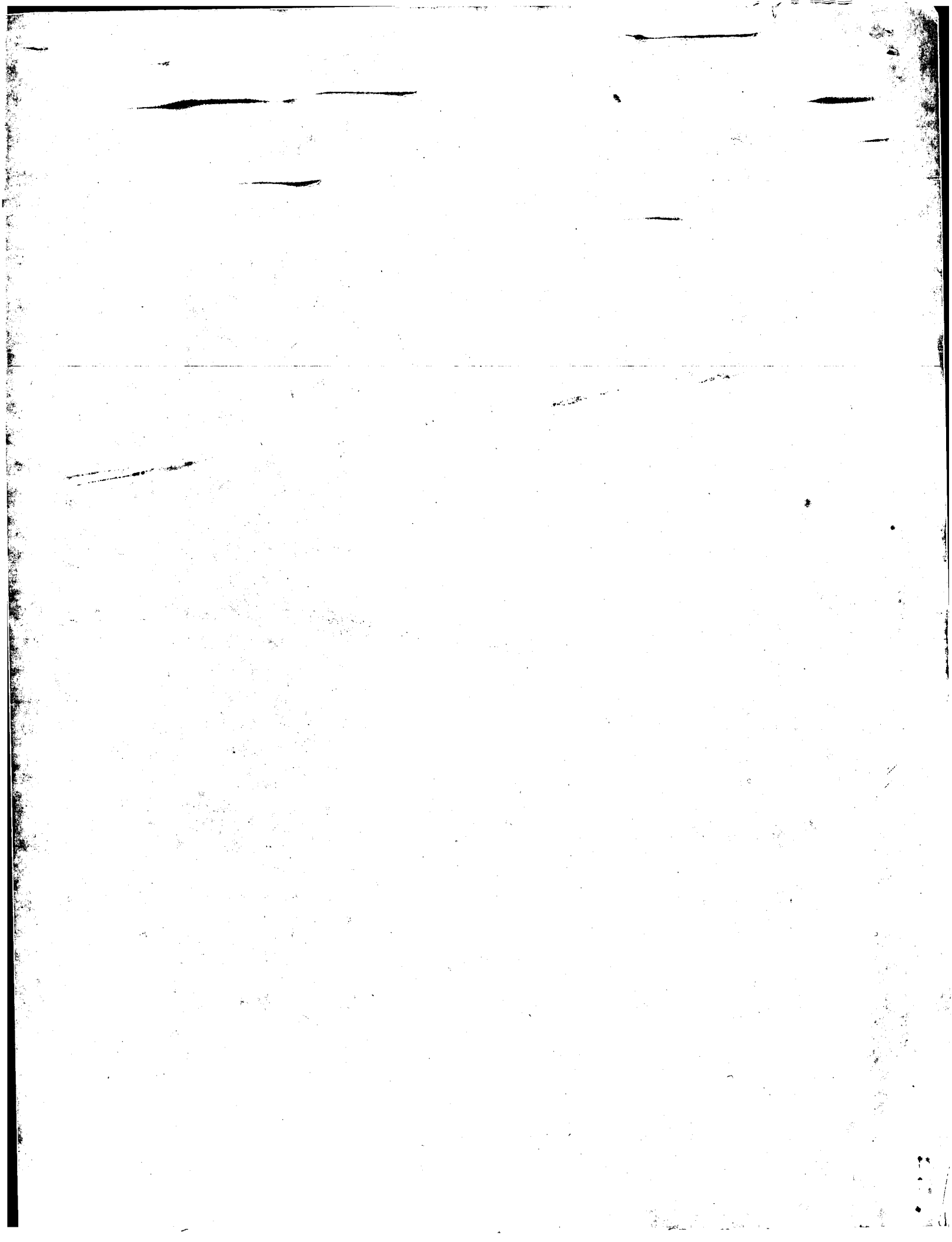
QY 181 FRKHDPATYSRCGSPWITPRCLVDYPYRLWHYPCTINVTIFKIRMYVGVGVEHRLAAACN 240
DB 586 FRKHPEATYTKCGSPWLTPRCLVDYPYRLWHYPCTVNETTKVRYMYGVGVEHRLVAAACN 645

QY 241 WTRGERCDLEDRDR 254
DB 646 WTRGERCNLEDRDR 659

Search completed: March 6, 2001, 11:55:12
Job time: 269 sec
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:50:43 ; Search time 75.33 Seconds  
(without alignments)  
261.396 Million cell updates/sec

Title: US-09-407-430-1

Perfect score: 921

Sequence: 1 MNSKGYPTQPTYPVQPGN.....VTQRKGNFPMGSDGGYTIW 168

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP-archaea.\*
- 2: sp-bacteria.\*
- 3: sp-fungi.\*
- 4: sp-human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp-organelle.\*
- 9: sp-phage.\*
- 10: sp-plant.\*
- 11: sp-rodent.\*
- 12: sp-virus.\*
- 13: sp-vertebrate.\*
- 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	921	100.0	168	4	Q15038
2	911	98.9	168	11	O88675
3	129	14.0	502	4	Q9UG75
4	126	13.7	856	4	O76022
5	124	13.5	262	3	O94274
6	124	13.5	291	10	Q9SKI0
7	124	13.5	494	10	Q93367
8	124	13.5	1840	10	Q9S2S8
9	123.5	13.4	2703	5	Q9VEG7
10	123.5	13.4	2715	5	O61603
11	123	13.4	625	10	Q99366
12	121.5	13.2	977	10	Q99368
13	120.5	13.1	468	5	Q9NF32
14	120.5	13.1	469	5	Q9WSD6
15	120	13.0	1126	5	Q19371
16	119	12.9	329	4	Q9NRQ2
17	119	12.9	1541	5	O15837
18	118.5	12.9	239	5	Q9W2R5
19	118	12.8	1690	5	O77165

20	117.5	12.8	530	10	O9LYK5
21	117.5	12.8	926	5	O9W3G1
22	117.5	12.8	1811	5	O9XZU8
23	117.5	12.8	1887	5	O9VYX6
24	117.5	12.8	1889	5	O9XZS2
25	117.5	12.8	1889	5	O9XZS1
26	117.5	12.8	1889	5	O9XZU9
27	117	12.7	1862	5	O20090
28	116.5	12.6	990	4	O9UG03
29	116.5	12.6	1345	4	O9P257
30	116.5	12.6	1553	5	O96452
31	116	12.6	261	4	O9P112
32	115	12.5	450	5	O9VYF9
33	114.5	12.4	440	5	O9VRL3
34	114.5	12.4	552	5	O76861
35	114.5	12.4	574	4	O08397
36	114	12.4	472	3	O59907
37	113.5	12.3	414	3	O94231
38	113.5	12.3	857	3	O9P7E8
39	113.5	12.3	1039	5	O9VKG4
40	113	12.3	261	11	P97765
41	113	12.3	306	11	O35449
42	113	12.3	1605	5	O96446
43	112	12.2	1668	4	O15026
44	112	12.2	2971	4	O9Y5L9
45	111.5	12.1	237	5	O17242

## ALIGNMENTS

RESULT 1

Q15038 PRELIMINARY: PRT; 168 AA.  
AC Q15038;  
DT 01-NOV-1996 (TRENBLREL. 01, Created)  
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)  
DE KIAA0058 PROTEIN.  
GN KIAA0058.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RX SEQUENCE FROM N.A.  
RP MEDLINE=96051398; PubMed=7584044;  
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,  
RA Seki N., Kavarabayashi Y., Ishikawa K., Tabata S.;  
RT "Prediction of the coding sequences of unidentified human genes. II.  
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 1:223-229(1994).  
DR ENGL; D31767; BAA06545.1;  
DR INTERPRO; IPR000515;  
DR PROSITE; PS00402; BPD-TRANSP\_INN\_MEMBR; UNKNOWN\_1.  
SQ SEQUENCE 168 AA; 17319 MW; 49F1B6D281E24AAC CRC64;

Query Match 100.0%; Score 921; DB 4; Length 168;  
Best Local Similarity 100.0%; Pred No. 6.2e-68;  
Matches 168; Conservative 0; Mismatches 0; Indels 0; Saps 0;  
QY 1 MNSKGYPTQPTYPVQPGNPVQPTLHLPAQPPYTDAPPAYSELYRPSFVHPGAATVTF 60  
Db 1 MNSKGYPTQPTYPVQPGNPVQPTLHLPAQPPYTDAPPAYSELYRPSFVHPGAATVTF 60  
QY 61 MSAAFPGASLYLPMASVAVGLPGSTIPMAYVPGVPIYPPGTVLVEGGYDAGARFGAGA 120  
Db 61 MSAAFPGASLYLPMASVAVGLPGSTIPMAYVPGVPIYPPGTVLVEGGYDAGARFGAGA 120  
QY 121 TAGNIPPPPGPPNAAQLAVMQANVLVTQRKGNFPMGSDGGYTIW 168  
|||||

Db 121 TAGNIPPPPPGCPNNAQLAVMOGANVLVTQRKGNFFMGSGGGYTIW 168

RESULT 2  
O88675 PRELIMINARY; PRT; 168 AA.  
AC O88675;  
DT 01-NOV-1998 (TRENBLrel. 08, Created)  
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE PROLINE-RICH PROTEIN.  
GN PRB OR BPRD.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD1; TISSUE=BRAIN;  
RA Yang W., Mansour S.L.;  
RT "A proline-rich protein expressed in mouse brain."  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF085348; AAC34594.1;  
DR MGD; MGI:1344344; Prtb.  
DR INTERPRO; IPR000515;  
DR PROSITE; PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN.1.  
SQ SEQUENCE 168 AA; 17288 MW; 49EDDAA29D8E344C CRC64;

Query Match 98.9%; Score 911; DB 11; Length 168;  
Best Local Similarity 98.8%; Pred. No. 4.1e-67;  
Matches 166; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MNSKGOYPTQPTYPVQPPGNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHPGAAATVPT 60  
DB 1 MNSKGOYPTQPTYPVQPPGNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHPGAAATVPT 60  
QY 61 MSAAPPASLYLPMASQAVAGPLSTPMAYYPVGPSTVLPVGGYDAGARFGAGA 120  
DB 61 MSAAPPASLYLPMASQAVAGPLSTPMAYYPVGPSTVLPVGGYDAGARFGAGA 120  
QY 121 TAGNIPPPPPGCPNNAQLAVMOGANVLVTQRKGNFFMGSGGGYTIW 168  
DB 121 TAGNIPPPPPGCPNNAQLAVMOGANVLVTQRKGNFFMGSGGGYTIW 168

RESULT 3  
Q9UG75 PRELIMINARY; PRT; 502 AA.  
AC Q9UG75;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)  
DE HYPOTHETICAL 55.5 KDA PROTEIN (FRAGMENT).  
GN DKFZP586D0920.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=UTERUS;  
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL050146; CAB43291.1;  
DR INTERPRO; IPR000216;  
DR INTERPRO; IPR002965;  
DR PRINTS; PR00239; RHODOPNTAIL.  
DR PRINTS; PR01217; PRICHEXTENSIN.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 502 AA; 55469 MW; C3ACF1854700F6E8 CRC64;

Query Match 14.0%; Score 129; DB 4; Length 502;  
Best Local Similarity 34.0%; Pred. No. 0.0049;  
Matches 50; Conservative 8; Mismatches 45; Indels 44; Gaps 10;

QY 2 NSKGOY---PTQPTYPVQPPGNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHPGAAAT- 57  
DB 329 NNRGSYNRAPOQQPPQPP-PPQPP-----PQPP---PPPSYSPARNP-----PGASTY 375  
QY 58 -----VPTMSA--AFPGASLYLPMASQAVAGPL-----GSTIPMAYYPVGPST 103  
DB 376 NKNSNIPGSSANTSTPTVSSYSPQPSYQPPYNOGGYSQGYTATPPPPPPPAYNYGS- 434  
QY 104 VLVEGGYDAGARFGAGATAGNIPPPPP 130  
DB 435 ---YGYN-----PAPYTPPPPP 449

RESULT 4  
O76022 PRELIMINARY; PRT; 856 AA.  
AC O76022;  
DT 01-NOV-1998 (TRENBLrel. 08, Created)  
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)  
DE E1B-55 KDA-ASSOCIATED PROTEIN.  
GN E1B-AP5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98406198; PubMed=9733834;  
RA Gabler S., Schuett H., Groitl P., Wolf H., Shenk T., Dobner T.;  
RT "E1B 55-kilodalton-associated protein: a cellular protein with RNA-binding activity implicated in nucleocytoplasmic transport of adenovirus and cellular mRNAs."  
RL J. Virol. 72:7960-7971(1998).  
DR EMBL; AJ007509; CAA07548.1;  
DR INTERPRO; IPR00107;  
DR INTERPRO; IPR003034;  
DR PFAM; PF00622; SPRY; 1.  
DR PFAM; PF02037; SAP; 1.  
SQ SEQUENCE 856 AA; 95809 MW; 937D6ACD1BD45DFF CRC64;

Query Match 13.7%; Score 126; DB 4; Length 856;  
Best Local Similarity 34.0%; Pred. No. 0.015;  
Matches 50; Conservative 8; Mismatches 45; Indels 44; Gaps 10;  
QY 2 NSKGOY---PTQPTYPVQPPGNVYPTLHLPOAPPYTDAPPAYSELYRPSFVHPGAAAT- 57  
DB 683 NNRGSYNRAPOQQPPQPP-PPQPP-----PQPP---PPPSYSPARNP-----PGASTY 729  
QY 58 -----VPTMSA--AFPGASLYLPMASQAVAGPL-----GSTIPMAYYPVGPST 103  
DB 730 NKNSNIPGSSANTSTPTVSSYSPQPSYQPPYNOGGYSQGYTATPPPPPPPAYNYGS- 788  
QY 104 VLVEGGYDAGARFGAGATAGNIPPPPP 130  
DB 789 ---YGYN-----PAPYTPPPPP 803

RESULT 5  
O94274 PRELIMINARY; PRT; 262 AA.  
AC O94274;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)  
DE PROLINE-RICH PROTEIN.  
GN SPBP8B7.26.



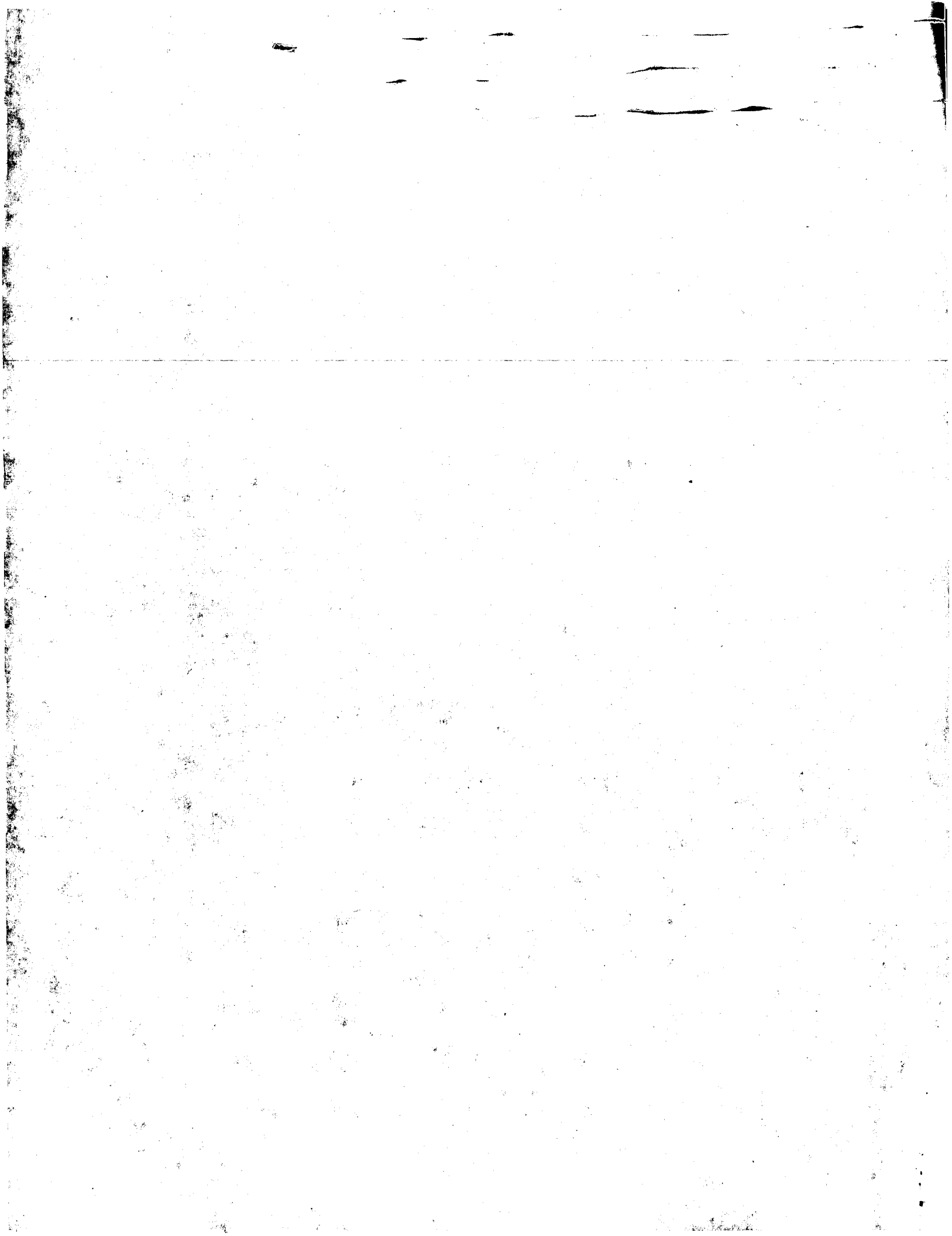
[illegible]











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:52:12 ; Search time 34.57 Seconds  
(without alignments)  
188.557 Million cell updates/sec

Title: US-09-407-430-2  
Perfect score: 2058  
Sequence: 1 ETHVTGGSAGHTVSGFVSL.....ADARVCSLWMLLSQAEA 363

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues  
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/PTOUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2058	100.0	480	1	US-08-440-103-14	Sequence 14, Appl
2	2058	100.0	480	1	US-08-440-542-14	Sequence 14, Appl
3	2058	100.0	480	1	US-08-231-368-14	Sequence 14, Appl
4	2058	100.0	480	1	US-08-440-210-14	Sequence 14, Appl
5	2058	100.0	663	3	US-08-824-057-3	Sequence 3, Appl
6	2058	100.0	2995	3	US-08-444-818-138	Sequence 138, Appl
7	2058	100.0	3011	1	US-08-440-103-36	Sequence 36, Appl
8	2058	100.0	3011	1	US-08-440-542-36	Sequence 36, Appl
9	2058	100.0	3011	1	US-07-910-760-10	Sequence 10, Appl
10	2058	100.0	3011	1	US-08-440-519-10	Sequence 10, Appl
11	2058	100.0	3011	1	US-08-231-368-36	Sequence 36, Appl
12	2058	100.0	3011	1	US-08-440-210-36	Sequence 36, Appl
13	2053	99.9	2772	3	US-08-444-818-89	Sequence 89, Appl
14	2053	99.8	2955	2	US-08-443-260-3	Sequence 3, Appl
15	2053	99.8	2955	3	US-08-442-805A-3	Sequence 3, Appl
16	2053	99.8	2955	3	US-08-443-900A-3	Sequence 3, Appl
17	2053	99.8	2955	3	US-08-444-818-124	Sequence 124, Appl
18	2053	99.8	3011	2	US-08-833-678A-6	Sequence 6, Appl
19	2053	99.8	3011	3	US-08-444-818-177	Sequence 177, Appl
20	2053	99.8	3011	4	PCT-US91-02225-10	Sequence 10, Appl
21	2037	99.0	738	4	PCT-US92-06965A-5	Sequence 5, Appl
22	2036	98.9	402	1	US-08-460-806-13	Sequence 13, Appl
23	2036	98.9	402	1	US-08-325-630-13	Sequence 13, Appl
24	2014	97.9	403	2	US-08-483-695-39	Sequence 39, Appl
25	2014	97.9	403	2	US-07-965-285-39	Sequence 39, Appl
26	2014	97.9	403	2	US-08-487-231-39	Sequence 39, Appl
27	1982	96.3	2894	2	US-08-466-975A-23	Sequence 23, Appl
28	1982	96.3	2894	2	US-08-391-671A-23	Sequence 23, Appl

29	1982	96.3	2894	3	US-08-467-902A-23	Sequence 23, Appl
30	1957	95.1	409	1	US-08-440-103-21	Sequence 21, Appl
31	1957	95.1	409	1	US-08-440-103-24	Sequence 24, Appl
32	1957	95.1	409	1	US-08-440-542-21	Sequence 21, Appl
33	1957	95.1	409	1	US-08-440-542-24	Sequence 24, Appl
34	1957	95.1	409	1	US-08-231-368-21	Sequence 21, Appl
35	1957	95.1	409	1	US-08-231-368-24	Sequence 24, Appl
36	1957	95.1	409	1	US-08-440-210-21	Sequence 21, Appl
37	1957	95.1	409	1	US-08-440-210-24	Sequence 24, Appl
38	1956	95.0	621	1	US-07-748-292-7	Sequence 7, Appl
39	1956	95.0	622	4	PCT-US92-06965A-4	Sequence 4, Appl
40	1955	95.0	3011	3	US-08-811-566-20	Sequence 20, Appl
41	1952	94.8	967	1	US-08-188-281B-13	Sequence 13, Appl
42	1952	94.8	967	4	PCT-US94-07280-13	Sequence 13, Appl
43	1952	94.8	967	4	PCT-US95-01087-13	Sequence 13, Appl
44	1952	94.8	1648	1	US-08-188-281B-12	Sequence 12, Appl
45	1952	94.8	1648	4	PCT-US94-07280-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-08-440-103-14  
; Sequence 14, Application US/08440103  
; Patent No. 5670152  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; APPLICANT: Houghton, Michael  
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,103  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 480 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-440-103-14

Query Match 100.0%; Score 2058; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.4e-196;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLNSALNCNDSLNTGWLGLFY 60  
Db 15 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLNSALNCNDSLNTGWLGLFY 74  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPYCWHYPKPGIIPAKSVC 120  
Db 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPYCWHYPKPGIIPAKSVC 134  
QY 121 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 135 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 194  
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240  
Db 195 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 254  
QY 241 YTFKIRMYGGVEHRLAEACNWTGRGCDLEDNRSELSPLLLTTTQWQVLPSCFTTLP 300  
Db 255 YTFKIRMYGGVEHRLAEACNWTGRGCDLEDNRSELSPLLLTTTQWQVLPSCFTTLP 314  
QY 301 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 360  
Db 315 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 374  
QY 361 AEA 363  
Db 375 AEA 377

## RESULT 2

US-08-440-542-14  
; Sequence 14, Application US/08440542  
; Patent No. 5670153  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; APPLICANT: Houghton, Michael  
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,542  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 480 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-440-542-14

Query Match 100.0%; Score 2058; DB 1: Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.4e-196;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLNSALNCNDSLNTGWLGLFY 60  
Db 15 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSHLNSALNCNDSLNTGWLGLFY 74  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPYCWHYPKPGIIPAKSVC 120  
Db 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPDRPYCWHYPKPGIIPAKSVC 134  
QY 121 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 135 GPVYCFPTSPVVVGTDRSGAPTSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKV 194  
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240  
Db 195 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 254  
QY 241 YTFKIRMYGGVEHRLAEACNWTGRGCDLEDNRSELSPLLLTTTQWQVLPSCFTTLP 300  
Db 255 YTFKIRMYGGVEHRLAEACNWTGRGCDLEDNRSELSPLLLTTTQWQVLPSCFTTLP 314  
QY 301 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 360  
Db 315 ALSTGLHLHQNIVDVOYLXGVSSIASWAIKWEYVVLFLLLADARVCSCLWMLLSQ 374  
QY 361 AEA 363  
Db 375 AEA 377

## RESULT 3

US-08-231-368-14  
; Sequence 14, Application US/08231368  
; Patent No. 5756312  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; APPLICANT: Houghton, Michael  
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 480 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-368-14

Query Match 100.0%; Score 2058; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.4e-196;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 60  
Db 15 ETHVTGGSAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 74  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCWHPKPGIVPAKSVC 120  
Db 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCWHPKPGIVPAKSVC 134  
QY 121 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 180  
Db 135 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 194  
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 240  
Db 195 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 254  
QY 241 YTFIKRMVYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQOVLPSCFTTLP 300  
Db 255 YTFIKRMVYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQOVLPSCFTTLP 314  
QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360  
Db 315 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 374  
QY 361 AEA 363  
Db 375 AEA 377

RESULT 4  
US-08-440-210-14  
; Sequence 14, Application US/08440210  
; Patent No. 5766845  
; GENERAL INFORMATION:  
; APPLICANT: Welner, Amy J.  
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE: 12-MAY-1995  
; APPLICATION NUMBER: US/08/440,210  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.

; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 480 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-440-210-14

Query Match 100.0%; Score 2058; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.4e-196;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 60  
Db 15 ETHVTGGSAGHTVSGFVSLAPGAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 74  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCWHPKPGIVPAKSVC 120  
Db 75 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQRPYCWHPKPGIVPAKSVC 134  
QY 121 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 180  
Db 135 GPVYCTPSPVVVGTTRDSGAPTSYSGENDTDVFLNTRPPLGNFPGCTWMNSTGFTKV 194  
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 240  
Db 195 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTIN 254  
QY 241 YTFIKRMVYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQOVLPSCFTTLP 300  
Db 255 YTFIKRMVYGVGVEHRLAECNWTGERCDLEDRDRSELSPLLLTTTQOVLPSCFTTLP 314  
QY 301 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360  
Db 315 ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 374  
QY 361 AEA 363  
Db 375 AEA 377

RESULT 5  
US-08-824-057-3  
; Sequence 3, Application US/08824057  
; Patent No. 6121020  
; GENERAL INFORMATION:  
; APPLICANT: SELBY, MARK  
; APPLICANT: HOUGHTON, MICHAEL  
; TITLE OF INVENTION: NOVEL HEPATITIS C EI AND E2 TRUNCATED  
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS OF OBTAINING THE SAME  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHIRON CORPORATION  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/824,057  
; FILING DATE:

CLASSIFICATION:  
; PRIOR APPLICATION NUMBER: US/08/282,959  
; FILING DATE: 29-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0987,001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 663 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-824-057-3

Query Match 100.0%; Score 2058; DB 3; Length 663;  
Best Local Similarity 100.0%; Pred. No. 3.7e-196;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 60  
DB 21 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 80  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYWHYPPKPGCIVPAKSVC 120  
DB 81 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYWHYPPKPGCIVPAKSVC 140  
QY 121 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 180  
DB 141 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 200  
QY 181 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240  
DB 201 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 260  
QY 241 YTIKIRMYGVGVEHRLAECNWTGRCDELDLDRSLSPLLLTTTQWQVLPSCFTTLP 300  
DB 261 YTIKIRMYGVGVEHRLAECNWTGRCDELDLDRSLSPLLLTTTQWQVLPSCFTTLP 320  
QY 301 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 360  
DB 321 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 380  
QY 361 AEA 363  
DB 381 AEA 383

RESULT 6  
US-08-444-818-138  
; Sequence 138, Application US/08444818  
; Patent No. 6150087  
; GENERAL INFORMATION:  
; APPLICANT: Chien, David Y.  
; APPLICANT: Rutter, William J.  
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines  
; NUMBER OF SEQUENCES: 777  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,818  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US/08/403,590  
FILING DATE: 14-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Harbin, Alisa A.  
REGISTRATION NUMBER: 33,895  
REFERENCE/DOCKET NUMBER: 0110,002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 359-3876  
TELEFAX: (508) 359-3885  
INFORMATION FOR SEQ ID NO: 138:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2995 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-444-818-138

Query Match 100.0%; Score 2058; DB 3; Length 2995;  
Best Local Similarity 100.0%; Pred. No. 3.2e-195;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 60  
DB 384 ETHVTGSGAGHTVSGFVSLAPGAKQNVQLINTNGSWHLNSTALNCNDSLTGWLGLFY 443  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYWHYPPKPGCIVPAKSVC 120  
DB 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDPQPCYWHYPPKPGCIVPAKSVC 503  
QY 121 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 180  
DB 504 GPVYCFTPSPVVGTTDRSGAPTSWGENDTDFVLNTRPLGNWFGCTWMNSTGFTKV 563  
QY 181 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 240  
DB 564 CGAPPCVTGGAGNNTLHCPDTCFRKHDPATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN 623  
QY 241 YTIKIRMYGVGVEHRLAECNWTGRCDELDLDRSLSPLLLTTTQWQVLPSCFTTLP 300  
DB 624 YTIKIRMYGVGVEHRLAECNWTGRCDELDLDRSLSPLLLTTTQWQVLPSCFTTLP 683  
QY 301 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 360  
DB 684 ALSTGLIHLHONIVDQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLNWMMLLSQ 743  
QY 361 AEA 363  
DB 744 AEA 746

RESULT 7  
US-08-440-103-36  
; Sequence 36, Application US/08440103  
; Patent No. 5670152  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; APPLICANT: Houghton, Michael  
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA



US-07-910-760-10  
; Sequence 10, Application US/07910760  
; Patent No. 5683864  
; GENERAL INFORMATION:  
; APPLICANT: Houghton, Michael  
; APPLICANT: Choo, Qui-Lim  
; APPLICANT: Kuo, George  
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: P.O. Box 8097 (Int. Prop. R-440)  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/910,760  
; APPLICATION NUMBER: US/07/910,760  
; FILING DATE: 07-JUL-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Blackburn Esq., Robert P.  
; REGISTRATION NUMBER: 30,447  
; REFERENCE/DOCKET NUMBER: 0101.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2702  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-910-760-10

Query Match 100.0%; Score 2058; DB 1; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 3.2e-195;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ETHVTGSGAGTSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 60  
Db 384 ETHVTGSGAGTSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 443  
Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDQPCYWHYPPKPGIYPAKSV 120  
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDQPCYWHYPPKPGIYPAKSV 503  
Qy 121 GPVYCFTSPVVVGTDRSGAPTSWGENDTDVFLNTRPPLGNWFCCTWMNSTGFTKV 180  
Db 504 GPVYCFTSPVVVGTDRSGAPTSWGENDTDVFLNTRPPLGNWFCCTWMNSTGFTKV 563  
Qy 181 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCIN 240  
Db 564 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCIN 623  
Qy 241 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWVLPSCFTTLP 300  
Db 624 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWVLPSCFTTLP 683  
Qy 301 ALSTGLIHLHQNIVDQVLYGVGSSIASWAKWEYVVLFLLLADARVCSCLMMMLLSQ 360  
Db 684 ALSTGLIHLHQNIVDQVLYGVGSSIASWAKWEYVVLFLLLADARVCSCLMMMLLSQ 743  
Qy 361 AEA 363  
Db 744 AEA 746

RESULT 10  
US-08-440-519-10  
; Sequence 10, Application US/08440519  
; Patent No. 5712087  
; GENERAL INFORMATION:  
; APPLICANT: Houghton, Michael  
; APPLICANT: Choo, Qui-Lim  
; APPLICANT: Kuo, George  
; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: P.O. Box 8097 (Int. Prop. R-440)  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94662-8097  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/440,519  
; APPLICATION NUMBER: US/08/440,519  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/910,760  
; FILING DATE: 07-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Blackburn Esq., Robert P.  
; REGISTRATION NUMBER: 30,447  
; REFERENCE/DOCKET NUMBER: 0101.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2702  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-440-519-10

Query Match 100.0%; Score 2058; DB 1; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 3.2e-195;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ETHVTGSGAGTSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 60  
Db 384 ETHVTGSGAGTSGFVSLAPGAKQNVQLINTNGSWHLNLTALNCNDSLNTGLAGLFY 443  
Qy 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDQPCYWHYPPKPGIYPAKSV 120  
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDQPCYWHYPPKPGIYPAKSV 503  
Qy 121 GPVYCFTSPVVVGTDRSGAPTSWGENDTDVFLNTRPPLGNWFCCTWMNSTGFTKV 180  
Db 504 GPVYCFTSPVVVGTDRSGAPTSWGENDTDVFLNTRPPLGNWFCCTWMNSTGFTKV 563  
Qy 181 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCIN 240  
Db 564 CGAPPCVIGGAGNNTLHCTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCIN 623  
Qy 241 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWVLPSCFTTLP 300  
Db 624 YTIKIRMYGVGVEHRLAECNWTGERCDLEDRLDRSELSPLLLTTTQWVLPSCFTTLP 683  
Qy 301 ALSTGLIHLHQNIVDQVLYGVGSSIASWAKWEYVVLFLLLADARVCSCLMMMLLSQ 360



Db 684 ALSTGLIHLHONIVDVQYLYGVGSSIASWALKWEYVVLFLLLADARVCSCLWMLLSQ 743  
QY 361 AEA 363  
Db 744 AEA 746  
RESULT 11  
US-08-231-368-36  
; Sequence 36, Application US/08231368  
; Patent No. 5756312  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; APPLICANT: Houghton, Michael  
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-368-36

Query Match 100.0%; Score 2058; DB 1; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 3.2e-195;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60  
Db 384 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 443  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVC 120  
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVC 503  
QY 121 GPVYCFTSPVVVGTTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 504 GPVYCFTSPVVVGTTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 563  
QY 181 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCPTIN 240  
Db 564 CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSCGSGPWITPRCLVDYPRYLWHYPCPTIN 623

QY 241 YTIKIRMYGGVEHRLLEAACNNTRGRCDEDRDRSELSPLLLTTTQWQVLPSCFTTLP 300  
Db 624 YTIKIRMYGGVEHRLLEAACNNTRGRCDEDRDRSELSPLLLTTTQWQVLPSCFTTLP 683  
QY 301 ALSTGLIHLHONIVDVQYLYGVGSSIASWALKWEYVVLFLLLADARVCSCLWMLLSQ 360  
Db 684 ALSTGLIHLHONIVDVQYLYGVGSSIASWALKWEYVVLFLLLADARVCSCLWMLLSQ 743  
QY 361 AEA 363  
Db 744 AEA 746  
RESULT 12  
US-08-440-210-36  
; Sequence 36, Application US/08440210  
; Patent No. 5766845  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; APPLICANT: Houghton, Michael  
; TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,210  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3011 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-440-210-36

Query Match 100.0%; Score 2058; DB 1; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 3.2e-195;  
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 60  
Db 384 ETHVTGGSAGHTVSGFVSLAPGAKONVOLINTNGSWHLNSTALNCNDSLNTGWLGLFY 443  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVC 120  
Db 444 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDQRPYCHWYPPKPGIIVPAKSVC 503  
QY 121 GPVYCFTSPVVVGTTDRSGAPTSYSGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180

Db 504 GPVYCFPSVVVGTDRSGAPYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 563  
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 240  
Db 564 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 623  
QY 241 YTIKIRMYGGVHRLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 300  
Db 624 YTIKIRMYGGVHRLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 683  
QY 301 ALSTGLHLHONIVDQYLYGVGSSIASWAIKWEYVYVLLFLLADARVCSCLNMMLLSQ 360  
Db 684 ALSTGLHLHONIVDQYLYGVGSSIASWAIKWEYVYVLLFLLADARVCSCLNMMLLSQ 743  
QY 361 AEA 363  
Db 744 AEA 746  
RESULT 13  
US-08-444-818-89  
; Sequence 89, Application US/08444818  
; Patent No. 6150087  
; GENERAL INFORMATION:  
; APPLICANT: Chien, David Y.  
; APPLICANT: Rutter, William J.  
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines  
; NUMBER OF SEQUENCES: 777  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/444.818  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/403.590  
; FILING DATE: 14-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harbin, Alisa A.  
; REGISTRATION NUMBER: 33,895  
; REFERENCE/DOCKET NUMBER: 0110.002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (508)359-3876  
; TELEFAX: (508)359-3885  
; INFORMATION FOR SEQ ID NO: 89:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2772 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-444-818-89  
Query Match 99.9%; Score 2055; DB 3; Length 2772;  
Best Local Similarity 99.7%; Pred. No. 5.7e-195;  
Matches 362; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ETHVTGSGAGHTVSGFVSLAPAKQNVQINTNGSWHLNLTALNCNDSLTNGWLAGLFY 60  
Db 270 ETHVTGSGAGHTVSGFVSLAPAKQNVQINTNGSWHLNLTALNCNDSLTNGWLAGLFY 329  
QY 61 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDORPYCWHYPKPGCIVPAKSYC 120

Db 330 HHKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGPDORPYCWHYPKPGCIVPAKSYC 389  
QY 121 GPVYCFPSVVVGTDRSGAPYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 180  
Db 390 GPVYCFPSVVVGTDRSGAPYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKV 449  
QY 181 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 240  
Db 450 CGAPPCVIGGAGNNTLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 509  
QY 241 YTIKIRMYGGVHRLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 300  
Db 510 YTIKIRMYGGVHRLHCPDTCFRKHPDATYSRGSGPWTIPRCLVDYPRYRLWHYPCNTIN 569  
QY 301 ALSTGLHLHONIVDQYLYGVGSSIASWAIKWEYVYVLLFLLADARVCSCLNMMLLSQ 360  
Db 570 ALSTGLHLHONIVDQYLYGVGSSIASWAIKWEYVYVLLFLLADARVCSCLNMMLLSQ 629  
QY 361 AEA 363  
Db 630 AEA 632  
RESULT 14  
US-08-443-260-3  
; Sequence 3, Application US/08443260  
; Patent No. 5942234  
; GENERAL INFORMATION:  
; APPLICANT: RALSTON, ROBERT O.  
; APPLICANT: MARCUS, FRANK  
; APPLICANT: THUDUM, KENT B.  
; APPLICANT: GERVAISE, BARBARA A.  
; APPLICANT: HALL, JOHN A.  
; TITLE OF INVENTION: HEPATITIS C VIRUS ASIALOGLYCOPROTEINS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/443.260  
; FILING DATE: 17-MAY-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HARBIN, ALISA A.  
; REGISTRATION NUMBER: 33,895  
; REFERENCE/DOCKET NUMBER: 0154.006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2955 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 9  
; OTHER INFORMATION: /note= "There is a heterogeneity at  
; OTHER INFORMATION: this location; Xaa = Arg or Lys"  
; FEATURE:  
; NAME/KEY: Modified-site

LOCATION: 11  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Asn or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 176  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Ile or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 334  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Met or Val"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 603  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Ile or Leu"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 848  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Asn or Tyr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1114  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Pro or Ser"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1117  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Ser or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1276  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Leu or Pro"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1454  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Cys or Tyr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1471  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Ser or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1877  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Glu or Gly"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1948  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - His or Leu"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1949  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Cys or Ser"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2021  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Gly or Val"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2349  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Ser or Thr"

FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2385  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Phe or Tyr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2386  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Ala or Ser"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2502  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Phe or Leu"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2690  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Gly or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2921  
OTHER INFORMATION: /note= "There is a heterogeneity at this location; Xaa - Arg or Gly"  
US-08-443-260-3

Query Match 99.8%; Score 2053; DB 2; Length 2955;

Best Local Similarity 99.7%; Pred No. 9.8e-195;

Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ETHVTGSGAGHTVSGFVSLAPCAKQNVOLINTNGSMHLNSTALNCNDSLNTGWLGLFY 60  
Db |||||  
QY 61 HKFNSSGCPERLASCRPLTDFDQGWGPISYANGSGDORPYCWHYPKPGIVPAKSYVC 120  
Db |||||  
QY 121 GPVYCFTPSPVVVGTTRDSGAPTSYMGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV 180  
Db |||||  
QY 181 CGAPCVIGGAGNNTLHCPTDFRKHDPDATYSCGSGPMITPRCLVDYDYPRLWHYPCTIN 240  
Db |||||  
QY 241 YTIKIRMYVGVGVEHRLAACHNWTGERCDLEDORSELSPLLLTTTQVQLPCSFITLP 300  
Db |||||  
QY 301 ALSTGLIHLHQNIVDVQVLYGVGSSIASWAIKWEYVVLFLLLADARVCSCLMMMLLSQ 360  
Db |||||  
QY 361 AEA 363  
Db |||||  
Db 744 AEA 746

RESULT 15

US-08-442-805A-3

; Sequence 3, Application US/08442805A

; Patent No. 6074846

; GENERAL INFORMATION:

; APPLICANT: RALSTON, ROBERT O.

; APPLICANT: MARCUS, FRANK

; APPLICANT: THUDLUM, KENT B.

; APPLICANT: GERVASE, BARBARA A.

; APPLICANT: HALL, JOHN A.

; TITLE OF INVENTION: HEPATITIS C VIRUS ASIALOGLYCOPROTEINS

; NUMBER OF SEQUENCES: 3

OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Leu or Pro"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	1454
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Cys or Tyr"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	1471
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Ser or Thr"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	1877
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Glu or Gly"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	1948
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = His or Leu"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	1949
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Cys or Ser"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2021
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Gly or Val"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2349
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Ser or Thr"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2385
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa -Phe or Tyr"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2386
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Ala or Ser"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2502
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Phe or Leu"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2690
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Gly or Arg"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	2921
OTHER INFORMATION:	/note= "There is a heterogeneity at
OTHER INFORMATION:	this location; Xaa = Arg or Gly"
OTHER INFORMATION:	442-805A-3

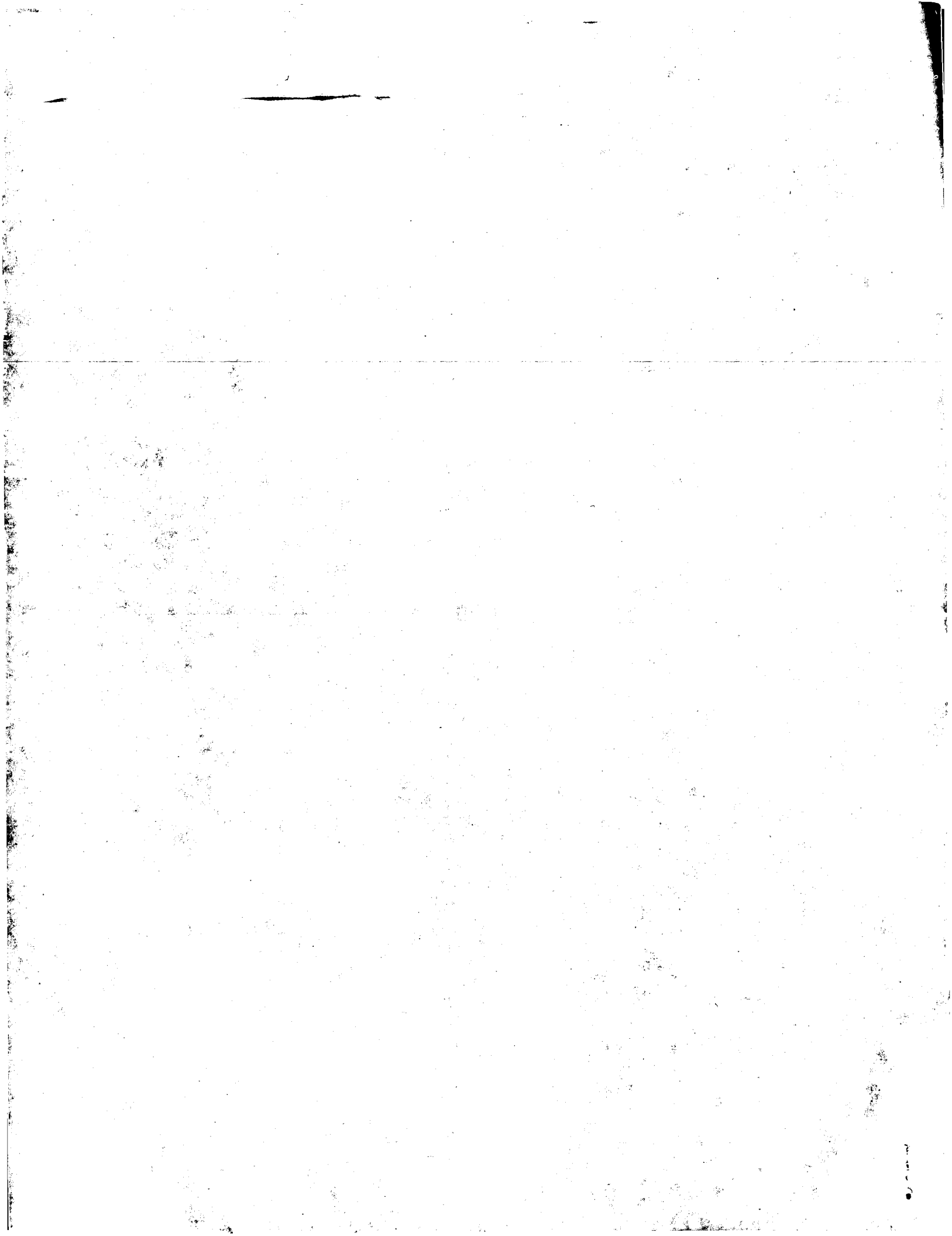
```

Query Match          99.8%; Score 2053; DB 3; Length 2955;
Best Local Similarity 99.7%; Pred. No. 9.8e-195;
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 ETHVTGGSAGHTVSGFVSLLPAGAKQVQLTNTGSHSLSTALNCNDSLTGWLGLCFY 60
384 ETHVTGGSAGHTVSGFVSLLPAGAKQVQLTNTGSHSLSTALNCNDSLTGWLGLCFY 443

```

QY	61	HHKFNSSGCCPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCWHPKPGIYPAKSVC	120
Db	444	HHKFNSSGCCPERLASCRPLTDFDQGWGPISYANGSGPDQRPYCWHPKPGIYPAKSVC	503
QY	121	GPVYCFTTSPVVVGTDRSGAPTYSNGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV	180
Db	504	GPVYCFTTSPVVVGTDRSGAPTYSNGENDTDVFLNNTRPPLGNWFGCTWMNSTGFTKV	563
QY	181	CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN	240
Db	564	CGAPPCVIGGAGNNTLHCPTDCFRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTIN	623
QY	241	YTIFKIRMYVGGVEHRLAECNNTRGERCLEDRDRSELSPLLLTTTQWQVLPSCFTTLP	300
Db	624	YTIFKIRMYVGGVEHRLAECNNTRGERCLEDRDRSELSPLLLTTTQWQVLPSCFTTLP	683
QY	301	ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLLFLLLADARVCSCLWMMLLSQ	360
Db	684	ALSTGLIHLHQNIVDVQYLYGVGSSIASWAIKWEYVVLLFLLLADARVCSCLWMMLLSQ	743
QY	361	AEA 363	
Db	744	AEA 746	

Search completed: March 6, 2001, 11:52:23  
Job time: 101 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 11:52:23 ; Search time 34.57 Seconds  
(without alignments)  
131.938 Million cell updates/sec

Title: US-09-407-430-3  
Perfect score: 1494  
Sequence: 1 GAKQNVQLINTNGSWHLNST.....LEAACNWTGRGCDLEDNRDR 254

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1494	100.0	402	1	US-08-460-806-13
2	1494	100.0	402	1	US-08-325-630-13
3	1494	100.0	403	2	US-08-483-693-39
4	1494	100.0	403	2	US-07-965-283-39
5	1494	100.0	403	2	US-08-487-231-39
6	1494	100.0	480	1	US-08-440-103-14
7	1494	100.0	480	1	US-08-440-542-14
8	1494	100.0	480	1	US-08-231-368-14
9	1494	100.0	480	1	US-08-440-210-14
10	1494	100.0	653	3	US-08-824-057-3
11	1494	100.0	2772	3	US-08-444-818-89
12	1494	100.0	2995	3	US-08-444-818-138
13	1494	100.0	3011	1	US-08-440-103-36
14	1494	100.0	3011	1	US-08-440-542-36
15	1494	100.0	3011	1	US-07-910-760-10
16	1494	100.0	3011	1	US-08-440-519-10
17	1494	100.0	3011	1	US-08-231-368-36
18	1494	100.0	3011	1	US-08-440-210-36
19	1489	99.7	2955	2	US-08-443-260-3
20	1489	99.7	2955	3	US-08-442-805A-3
21	1489	99.7	2955	3	US-08-443-900A-3
22	1489	99.7	2955	3	US-08-444-818-124
23	1489	99.7	3011	2	US-08-833-678A-6
24	1489	99.7	3011	3	US-08-444-818-177
25	1489	99.7	3011	4	PCT-US91-02225-10
26	1473	98.6	621	1	US-07-748-292-7
27	1473	98.6	622	4	PCT-US92-06965A-4
28	1473	98.6	738	4	PCT-US92-06965A-5

29	1473	98.6	2894	2	US-08-466-975A-23	Sequence 23, Appl
30	1473	98.6	2894	2	US-08-391-671A-23	Sequence 23, Appl
31	1473	98.6	2894	3	US-08-467-902A-23	Sequence 23, Appl
32	1426	95.4	409	1	US-08-440-103-21	Sequence 21, Appl
33	1426	95.4	409	1	US-08-440-103-24	Sequence 24, Appl
34	1426	95.4	409	1	US-08-440-542-21	Sequence 21, Appl
35	1426	95.4	409	1	US-08-231-368-21	Sequence 21, Appl
36	1426	95.4	409	1	US-08-440-542-24	Sequence 24, Appl
37	1426	95.4	409	1	US-08-231-368-24	Sequence 21, Appl
38	1426	95.4	409	1	US-08-440-210-21	Sequence 21, Appl
39	1426	95.4	409	1	US-08-440-210-24	Sequence 21, Appl
40	1418	94.9	3011	3	US-09-014-416-1	Sequence 1, Appl
41	1418	94.9	3012	3	US-08-811-566-2	Sequence 2, Appl
42	1416	94.8	3011	3	US-08-811-566-20	Sequence 20, Appl
43	1414	94.6	402	1	US-08-460-806-17	Sequence 17, Appl
44	1414	94.6	402	1	US-08-325-630-17	Sequence 17, Appl
45	1413	94.6	305	3	US-08-478-073-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-460-806-13  
; Sequence 13, Application US/08460806  
; Patent No. 5747241  
; GENERAL INFORMATION:  
; APPLICANT: MIYAMURA, TATSUO  
; APPLICANT: SAITO, IZUMU  
; APPLICANT: HARADA, SHIZUKO  
; APPLICANT: HONDA, YOSHIKAZU  
; TITLE OF INVENTION: DIAGNOSTIC REAGENT FOR HEPATITIS C  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08460.806  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325.630  
; FILING DATE: 19-OCT-1994  
; APPLICATION NUMBER: US 07/956.993  
; FILING DATE: 06-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5747241man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 4667-001-0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 402 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-460-806-13

Query Match 100.0% Score 1494; DB 1; Length 402;

Best Local Similarity	100.0%;	Pred. No. 4e-141;
Matches 254;	Conservative	0; Mismatches 0;
		Indels 0; Gaps 0;

1	GAKQVQLINTNGSHLNSTALNCNDSLNTGWLGLFVHHKFNSSGCPERLASCRPLTDF	60
yy		
bb		
67	GAKQVQLINTNGSHLNSTALNCNDSLNTGWLGLFVHHKFNSSGCPERLASCRPLTDF	126
61	DQGWGISYANGSGDQRPYCHWHPKPGIVPAKSVCGVPYCFTSPVVVGTTDRSGAP	120
yy		
bb		
127	DQGWGISYANGSGDQRPYCHWHPKPGIVPAKSVCGVPYCFTSPVVVGTTDRSGAP	186
121	TYSWGENDTDVFVLNTRPPLGNWFECCTWMNSTGTFTKVCAGPCVLTGGAGNNTLHCPTDC	180
yy		
bb		
187	TYSWGENDTDVFVLNTRPPLGNWFECCTWMNSTGTFTKVCAGPCVLTGGAGNNTLHCPTDC	246
181	FRKHDPATYSRCGSGPWITPRCLVDYPYLWHYPCTINITYIFKIRMYVGVGEHRLEAACN	240
yy		
bb		
247	FRKHDPATYSRCGSGPWITPRCLVDYPYLWHYPCTINITYIFKIRMYVGVGEHRLEAACN	306
241	WTRGERCDLEDNRD	254
yy		
bb		
307	WTRGERCDLEDNRD	320
bb		

```

3
US-08-483-695-39
; Sequence 39, Application US/08483695
; Patent No. 5866139
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,695
; FILING DATE:

```

Query Match 100.0%; Score 1494; DB 1; Length 402;



Query Match 100.0%; Score 1494; DB 2; Length 403;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 60  
DB 73 KAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 132

QY 61 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120  
DB 133 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 192

QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 180  
DB 193 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 252

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 240  
DB 253 FRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 312

QY 241 WTRGERCDLEDRDR 254  
DB 313 WTRGERCDLEDRDR 326

RESULT 4  
US-07-965-285-39  
; Sequence 39, Application US/07965285  
; Patent No. 5879904  
; GENERAL INFORMATION:  
; APPLICANT: Brechot, Christian  
; APPLICANT: Kremsdorf, Dina  
; APPLICANT: Porchon, Colette  
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a  
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic  
; TITLE OF INVENTION: Applications  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/965,285  
; FILING DATE: 18-MAR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 91 06 882  
; FILING DATE: 06-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 05286-0001-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 403 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-07-965-285-39

Query Match 100.0%; Score 1494; DB 2; Length 403;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 60  
DB 73 KAKONVOLINTNGSHLNTALNCNDSLNTGWLGLFYHHKFNSSGCGPERLASCRPLTDF 132

QY 61 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120  
DB 133 DOGWGPISYANGSGDPORPCWHPKPCGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 192

QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 180  
DB 193 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCAGAPPCVIGGAGNNTLHCPTDC 252

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 240  
DB 253 FRKHPDATYSRCGSGPWITPRCLVDYPYRLWHYPCTINTYFIKIRMYVGGVEHRLAECN 312

QY 241 WTRGERCDLEDRDR 254  
DB 313 WTRGERCDLEDRDR 326

RESULT 5  
US-08-487-231-39  
; Sequence 39, Application US/08487231  
; Patent No. 5919454  
; GENERAL INFORMATION:  
; APPLICANT: Brechot, Christian  
; APPLICANT: Kremsdorf, Dina  
; APPLICANT: Porchon, Colette  
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a  
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic  
; TITLE OF INVENTION: Applications  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,231  
; FILING DATE: 07-JUNE-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/965,285  
; FILING DATE: 18-MAR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 91 06 882  
; FILING DATE: 06-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meyers, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 05286-0001-02000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 403 amino acids  
; TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-487-231-39

Query Match 100.0%; Score 1494; DB 2; Length 403;  
Best Local Similarity 100.0%; Pred. No. 4e-141;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKNQVLINTNGSHLNTALNCNDSINTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
DB 73 GAKNQVLINTNGSHLNTALNCNDSINTGWLGLFYHHKFNSSGCPERLASCRPLTDF 132  
QY 61 DQGWPISYANGSGDPQRPYCHWHPKPGCGIVPAKSVCGPVYCTPTSPVVVGTDRSGAP 120  
DB 133 DQGWPISYANGSGDPQRPYCHWHPKPGCGIVPAKSVCGPVYCTPTSPVVVGTDRSGAP 192  
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180  
DB 193 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 252  
QY 181 FRKHDPATYSRGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGGVEHRLAECN 240  
DB 253 FRKHDPATYSRGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGGVEHRLAECN 312  
QY 241 WTRGERCDLEDRDR 254  
DB 313 WTRGERCDLEDRDR 326

RESULT 6

US-08-440-103-14  
Sequence 14, Application US/08440103  
Patent No. 5670152  
GENERAL INFORMATION:  
APPLICANT: Weiner, Amy J.  
APPLICANT: Houghton, Michael  
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,103  
FILING DATE: 12-MAY-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,368  
FILING DATE:  
APPLICATION NUMBER: US 07/759,575  
FILING DATE: 13-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: McClung, Barbara G.  
REGISTRATION NUMBER: 33,113  
REFERENCE/DOCKET NUMBER: 0205.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2708  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-440-103-14

Query Match 100.0%; Score 1494; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 5e-141;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKNQVLINTNGSHLNTALNCNDSINTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60  
DB 37 GAKNQVLINTNGSHLNTALNCNDSINTGWLGLFYHHKFNSSGCPERLASCRPLTDF 96  
QY 61 DQGWPISYANGSGDPQRPYCHWHPKPGCGIVPAKSVCGPVYCTPTSPVVVGTDRSGAP 120  
DB 97 DQGWPISYANGSGDPQRPYCHWHPKPGCGIVPAKSVCGPVYCTPTSPVVVGTDRSGAP 156  
QY 121 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180  
DB 157 TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 216  
QY 181 FRKHDPATYSRGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGGVEHRLAECN 240  
DB 217 FRKHDPATYSRGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGGVEHRLAECN 276  
QY 241 WTRGERCDLEDRDR 254  
DB 277 WTRGERCDLEDRDR 290

RESULT 7

US-08-440-542-14  
Sequence 14, Application US/08440542  
Patent No. 5670153  
GENERAL INFORMATION:  
APPLICANT: Weiner, Amy J.  
APPLICANT: Houghton, Michael  
TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,542  
FILING DATE: 12-MAY-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/231,368  
FILING DATE:  
APPLICATION NUMBER: US 07/759,575  
FILING DATE: 13-SEP-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: McClung, Barbara G.  
REGISTRATION NUMBER: 33,113  
REFERENCE/DOCKET NUMBER: 0205.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2708  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-440-542-14

Query Match 100.0%; Score 1494; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 5e-141;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQINTNGSHLNTALNCNDSLTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 60  
DB 37 GAKQNVQINTNGSHLNTALNCNDSLTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 96

QY 61 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120  
DB 97 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 156

QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180  
DB 157 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 216

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGVGVEHRLAECN 240  
DB 217 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGVGVEHRLAECN 276

QY 241 WTRGERCDLEDRDR 254  
DB 277 WTRGERCDLEDRDR 290

RESULT 8  
US-08-231-368-14  
; Sequence 14, Application US/08231368  
; Patent No. 5756312  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 480 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-368-14

Query Match 100.0%; Score 1494; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 5e-141;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAKQNVQINTNGSHLNTALNCNDSLTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 60  
DB 37 GAKQNVQINTNGSHLNTALNCNDSLTGWLGLFVHHKFNSSGCGPERLASCRPLTDF 96

QY 61 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 120  
DB 97 DQGWGPTSYANGSGPDQRPYCHWYPPKPCGIVPAKSVCGPVYCFPTSPVVVGTDRSGAP 156

QY 121 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 180  
DB 157 TYSWGENDTDFVLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC 216

QY 181 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGVGVEHRLAECN 240  
DB 217 FRKHPDATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYTIKIRMYVGVGVEHRLAECN 276

QY 241 WTRGERCDLEDRDR 254  
DB 277 WTRGERCDLEDRDR 290

RESULT 9  
US-08-440-210-14  
; Sequence 14, Application US/08440210  
; Patent No. 5766845  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, Amy J.  
; TITLE OF INVENTION: Immunoactive Polypeptide Compositions  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,210  
; FILING DATE: 12-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,368  
; FILING DATE:  
; APPLICATION NUMBER: US 07/759,575  
; FILING DATE: 13-SEP-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McClung, Barbara G.  
; REGISTRATION NUMBER: 33,113  
; REFERENCE/DOCKET NUMBER: 0205.001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 601-2708  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 480 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-440-210-14

Query Match 100.0%; Score 1494; DB 1; Length 480;

```
Best Local Similarity 100.0%; Pred. No. 5e-141; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;

QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 37 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 96
QY 61 DQGWGPISYANGSGDPQRPYCWHPKPGIIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120
DB 97 DQGWGPISYANGSGDPQRPYCWHPKPGIIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 156
QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
DB 157 TYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 216
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYIFKIRMYVGVGVEHRLAECN 240
DB 217 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYIFKIRMYVGVGVEHRLAECN 276
QY 241 WTRGERCDLEDRDR 254
DB 277 WTRGERCDLEDRDR 290

RESULT 10
US-08-824-057-3
; Sequence 3, Application US/08824057
; Patent No. 6121020
; GENERAL INFORMATION:
; APPLICANT: SELBY, MARK
; APPLICANT: HOUGHTON, MICHAEL
; TITLE OF INVENTION: NOVEL HEPATITIS C E1 AND E2 TRUNCATED
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS OF OBTAINING THE SAME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,057
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,959
; FILING DATE: 29-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0987.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-824-057-3

Query Match 100.0%; Score 1494; DB 3; Length 663;
Best Local Similarity 100.0%; Pred. No. 7.6e-141; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;
```

```
QY 1 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 60
DB 43 GAKQNVQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSGCPERLASCRPLTDF 102
QY 61 DQGWGPISYANGSGDPQRPYCWHPKPGIIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 120
DB 103 DQGWGPISYANGSGDPQRPYCWHPKPGIIVPAKSVCGPVYCFPTSPVVVGTTRDSGAP 162
QY 121 TYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 180
DB 163 TYSWGENDTDFVNLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGAGNNTLHCPTDC 222
QY 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYIFKIRMYVGVGVEHRLAECN 240
DB 223 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCTINTYIFKIRMYVGVGVEHRLAECN 282
QY 241 WTRGERCDLEDRDR 254
DB 283 WTRGERCDLEDRDR 296

RESULT 11
US-08-444-818-89
; Sequence 89, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisha A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-89

Query Match 100.0%; Score 1494; DB 3; Length 2772;
Best Local Similarity 100.0%; Pred. No. 5e-140; Indels 0; Gaps 0;
Matches 254; Conservative 0; Mismatches 0;
```

Qy	61	DQNGPISYANGSPDQRPYCHWYPPKPCGIVPAKSVCGPVYCFTPSPVVVGTTRDSGAP	120
Db	352	DQNGPISYANGSPDQRPYCHWYPPKPCGIVPAKSVCGPVYCFTPSPVVVGTTRDSGAP	411
Qy	121	TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC	180
Db	412	TYSWGENDTDFVLNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVIGGAGNNTLHCPTDC	471
Qy	181	FRKHDPATYSCRSGGPWITPRCLVDYPIRLWHYPCTINTYIFKIRMYGVGVEHRLAECN	240
Db	472	FRKHDPATYSCRSGGPWITPRCLVDYPIRLWHYPCTINTYIFKIRMYGVGVEHRLAECN	531
Qy	241	WTRGERCDLEDNRD	254
Db	532	WTRGERCDLEDNRD	545

```

RESULT 12
US-08-444-818-138
; Sequence 138, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisha A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3895
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2995 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-444-818-138

```

	Query Match	100.0%	Score 1494;	DB 3;	Length 2995;
	Best Local Similarity	100.0%	Pred. No. 5,5e-140;		
	Matches 254;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GAKQWQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSCCPERLASCRPLTDF	60		
Db	406	GAKQWQLINTNGSWHLNSTALNCNDSLNTGWLGLFYHHKFNSSCCPERLASCRPLTDF	465		
QY	61	DGCGPISVANGSGDQRPYCWYHPKPCGIVPAKSVCGPVCFTSPVVGTTDRSGAP	120		
Db	466	DGCGPISVANGSGDQRPYCWYHPKPCGIVPAKSVCGPVCFTSPVVGTTDRSGAP	525		

```

QY 121 TYSWGEDTDVFLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVITGGAGNNTLHCPTDC 180
      |||||||
DB 526 TYSWGEDTDVFLNNTRPPLGNWFGCTWMNSTGFTKVCGAPPCVITGGAGNNTLHCPTDC 585
      |||||||
QY 181 FRKHPDATYSCGSGPWITPCLVDYDYPRLWHHPCTINTYTFIKIRMYVGVGVEHRLSAACN 240
      |||||||
DB 586 FRKHPDATYSCGSGPWITPCLVDYDYPRLWHHPCTINTYTFIKIRMYVGVGVEHRLSAACN 645
      |||||||
QY 241 WTRGERCDLEDRDR 254
      |||||||
DB 646 WTRGERCDLEDRDR 659
      |||||||

RESULT 13
US-08-440-103-36
; Sequence 36, Application US/08440103
; Patent No. 5670152
; GENERAL INFORMATION:
; APPLICANT: Weiner, Amy J.
; APPLICANT: Houghton, Michael
; TITLE OF INVENTION: Immunoactive polypeptide Compositions
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,103
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,368
; FILING DATE:
; APPLICATION NUMBER: US 07/759,575
; FILING DATE: 13-SEP-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0205.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2708
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-440-103-36

```

	Query Match	100.0%	Score 1494;	DB 1;	Length 3011;
	Best Local Similarity	100.0%	Pred. No. 5	6e-140;	
	Matches 254;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GAKONVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFVHHKNSGCGPERLASCRPLTDF	60		
Db	406	GAKONVOLINTNGSWHLNSTALNCNDSLNTGWLAGLFVHHKNSGCGPERLASCRPLTDF	60		
QY	61	DOGWGPIISYANGSGPDORPCYWHYPFRPCGIVTAPKASVCGGPGVYCFPTSPVVVGTTRDSGAP	120		
Db	466	DOGWGPIISYANGSGPDORPCYWHYPFRPCGIVTAPKASVCGGPGVYCFPTSPVVVGTTRDSGAP	525		

Qy 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 180  
Db 526 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 585  
Qy 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 240  
Db 586 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 645  
Qy 241 WTRGERCDLEDRDR 254  
Db 646 WTRGERCDLEDRDR 659

RESULT 14  
US-08-440-542-36  
: Sequence 36, Application US/08440542  
: Patent No. 5670153  
: GENERAL INFORMATION:  
: APPLICANT: Weiner, Amy J.  
: APPLICANT: Houghton, Michael  
: TITLE OF INVENTION: Immunoreactive Polypeptide Compositions  
: NUMBER OF SEQUENCES: 45  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Chiron Corporation  
: STREET: 4560 Horton Street  
: CITY: Emeryville  
: STATE: CA  
: COUNTRY: USA  
: ZIP: 94608  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/440.542  
: FILING DATE: 12-MAY-1995  
: CLASSIFICATION: 424  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/08/231.368  
: FILING DATE:  
: APPLICATION NUMBER: US 07/759,575  
: FILING DATE: 13-SEP-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: McClung, Barbara G.  
: REGISTRATION NUMBER: 33,113  
: REFERENCE/DOCKET NUMBER: 0205.001  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (510) 601-2708  
: TELEFAX: (510) 655-3542  
: INFORMATION FOR SEQ ID NO: 36:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3011 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
US-08-440-542-36

Query Match 100.0%; Score 1494; DB 1; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 5.6e-140;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFHHKFNSSGCPERLASCRPLTDF 60  
Db 406 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFHHKFNSSGCPERLASCRPLTDF 465  
Qy 61 DQGWGPISYANGSGPDQRPYCWHPKPGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120  
Db 466 DQGWGPISYANGSGPDQRPYCWHPKPGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 525  
Qy 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 180

Db 526 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 585  
Qy 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 240  
Db 586 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 645  
Qy 241 WTRGERCDLEDRDR 254  
Db 646 WTRGERCDLEDRDR 659

RESULT 15  
US-07-910-760-10  
: Sequence 10, Application US/07910760  
: Patent No. 5683864  
: GENERAL INFORMATION:  
: APPLICANT: Houghton, Michael  
: APPLICANT: Choo, Qui-Lim  
: TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
: NUMBER OF SEQUENCES: 12  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Chiron Corporation  
: STREET: P.O. Box 8097 (Int. Prop. R-440)  
: CITY: Emeryville  
: STATE: CA  
: COUNTRY: U.S.A.  
: ZIP: 94662-8097  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/07/910,760  
: FILING DATE: 07-JUL-1992  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Blackburn Esq., Robert P.  
: REGISTRATION NUMBER: 30,447  
: REFERENCE/DOCKET NUMBER: 0101.002  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (510) 601-2702  
: TELEFAX: (510) 655-3542  
: INFORMATION FOR SEQ ID NO: 10:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 3011 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
US-07-910-760-10

Query Match 100.0%; Score 1494; DB 1; Length 3011;  
Best Local Similarity 100.0%; Pred. No. 5.6e-140;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFHHKFNSSGCPERLASCRPLTDF 60  
Db 406 GAKONVOLINTNGSWHLNSTALNCNDSLTGWLGLFHHKFNSSGCPERLASCRPLTDF 465  
Qy 61 DQGWGPISYANGSGPDQRPYCWHPKPGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 120  
Db 466 DQGWGPISYANGSGPDQRPYCWHPKPGIVPAKSVCGPVYCFTPSPVVVGTDRSGAP 525  
Qy 121 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 180  
Db 526 TYSWGENDTDFVLLNTRPPLGNWFGCTWMNSTGFTKVCAPPVIGGAGNNTLHCPTDC 585  
Qy 181 FRKHDPATYSRCGSGPWITPRCLVDYPRYLWHYPCITNTYIFKIRMYVGGVEHRLAECN 240

Db 586 FRKHPDATYSRCGSPWITPRCLVDYDPRLWHYPCTINYTIKIRMYVVGVEHLEAACN 645

Qy 241 WTRGRCGLEDRDR 254

Db 646 WTRGRCGLEDRDR 659

Search completed: March 6, 2001, 11:52:27  
Job time: 105 sec

